January 1996

Multi-band Oversampled Noise Shaping Analog to Digital Conversion

Pervez M. Aziz
University of Pennsylvania

Multi-band Oversampled Noise Shaping Analog to Digital Conversion

Abstract
Oversampled noise shaping analog to digital (A/D) converters, which are commonly known as delta-sigma ($\Delta\Sigma$) converters, have the ability to convert relatively low bandwidth signals with very high resolution. Such converters achieve their high resolution by oversampling, as well as processing the signal and quantization noise with different transfer functions. The signal transfer function (STF) is typically a delay over the signal band while the noise transfer function (NTF) is designed to attenuate quantization noise in the signal band. A side effect of the NTF is an amplification of the noise outside the signal band. Thus, a digital filter subsequently attenuates the out-of-band quantization noise.

The focus of this thesis is the investigation of $\Delta\Sigma$ architectures that increase the bandwidth where high resolution conversion can be achieved. It uses parallel architectures exploiting frequency or time slicing to meet this objective. Frequency slicing involves quantizing different portions of the signal frequency spectrum using several quantizers in parallel and then combining the results of the quantizers to form an overall result. Time slicing involves quantizing various groups of time domain signal samples with different quantizers in parallel and then combining the results of the quantizers to form an overall output.

Several interesting observations can be made from this general perspective of frequency and time slicing. Although the representation of a signal are completely equivalent in time or frequency, the thesis shows that this is not the case for known frequency and time sliced A/D architectures. The performance of such systems under ideal conditions are compared for PCM as well as for $\Delta\Sigma$ A/D converters. A multi-band frequency sliced architecture for delta-sigma conversion is proposed and its performance is included in the above comparison. The architecture uses modulators which realize different NTFs for different portions of the signal band. Each band is converted in parallel. A bank of FIR filters attenuates the out of-band noise for each band and achieves perfect reconstruction of the signal component. A design procedure is provided for the design of the filter bank with reduced computational complexity. The use of *complex* NTFs in the multi-band $\Delta\Sigma$ architecture is also proposed. The peformance of real and complex NTFs is compared. Performance evaluations are made for ideal systems as well as systems suffering from circuit implementation imperfections such as finite opamp gain and mismatched capacitor ratios.

Keywords
engineering, electrical

Multi-band Oversampled Noise Shaping Analog to Digital Conversion

Pervez Mirza Aziz

A DISSERTATION
in
Electrical Engineering

Presented to the Faculties of the
University of Pennsylvania in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

1996

Supervisors of Dissertation

Dr. Jan Van der Spiegel, Dr. Henrik V. Sorensen

Graduate Group Chairperson

Dr. Nader Engheta

Acknowledgements

I would like to thank those who have made this work possible. Special thanks are due to my advisors Drs. Van der Spiegel and Sorensen for their continual support, encouragement, guidance, and laborious reading of this thesis as well as thesis related publication drafts. Dr. Sorensen is warmly thanked for countless enthusiastic discussions on the digital signal processing aspects of the thesis.

Dr. T.R. Viswanathan, formerly of AT&T Bell Laboratories and now at Texas Instruments Research and Dr. Kenneth Laker are thanked for their encouragement and for being instrumental in initiating this research. Thanks are due to the Department of Electrical Engineering, University of Pennsylvania and former and present Graduate Chairs Drs. Sohrab Rabii and Nader Engheta for coordinating financial support through research and teaching fellowships. I would like to thank Drs. Saleem Kassam, Sumit Roy and Hugh Williams for taking the time to serve on my proposal defense committee. The same thanks is due to my dissertation defense committee members: Dr. Krishnaswamy Nagaraj formerly of AT&T Bell Laboratories and now at Texas Instruments, Drs. Magda El Zarki, Saleem Kassam, Kenneth Laker of the Electrical Engineering deparment.

Thanks are also due to Renee Kulba who has provided excellent administrative support for all aspects of my graduate education.

Steve Jantzi and Chris Ouslis of the University of Toronto are kindly thanked for their help in providing access to the software filtorX and for their very helpful suggestions on using it.

At the personal level I am pleased to acknowledge my wife Akila for her unwavering love and support, for persevering through many lonely evenings and weekends, and for packing many delicious lunches and dinners. Our son Tahir Mirza Aziz also deserves acknowledgement for providing the extra motivation needed to finish the pre-defense draft of this thesis which was completed three weeks before his birth (the thesis defense took place ten days before his birth). I would like to also thank my parents Dr. and Mrs. Azizul Islam for their encouragement. Finally, I need to thank my friend and colleague Dr. Necip Sayiner for his continual encouragement.

Abstract

Multi-band Oversampled Noise Shaping
Analog to Digital Conversion
Author: Pervez Mirza Aziz
Supervisors: Henrik V. Sorensen, Jan Van der Spiegel

Oversampled noise shaping analog to digital (A/D) converters, which are commonly known as delta-sigma ($\Delta\Sigma$) converters, have the ability to convert relatively low bandwidth signals with very high resolution. Such converters achieve their high resolution by oversampling, as well as processing the signal and quantization noise with different transfer functions. The signal transfer function (STF) is typically a delay over the signal band while the noise transfer function (NTF) is designed to attenuate quantization noise in the signal band. A side effect of the NTF is an amplification of the noise outside the signal band. Thus, a digital filter subsequently attenuates the out-of-band quantization noise.

The focus of this thesis is the investigation of $\Delta\Sigma$ architectures that increase the bandwidth where high resolution conversion can be achieved. It uses parallel architectures exploiting frequency or time slicing to meet this objective. Frequency slicing involves quantizing different portions of the signal frequency spectrum using several quantizers in parallel and then combining the results of the quantizers to form an overall result. Time slicing involves quantizing various groups of time domain signal samples with different quantizers in parallel and then combining the results of the quantizers to form an overall output.

Several interesting observations can be made from this general perspective of frequency and time slicing. Although the representation of a signal are completely equivalent in time or frequency, the thesis shows that this is not the case for known frequency and time sliced A/D architectures. The performance of such systems under ideal conditions are compared for PCM as well as for $\Delta\Sigma$ A/D converters. A multi-band frequency sliced architecture for delta-sigma conversion is proposed and its performance is included in the above comparison. The architecture uses modulators which realize different NTFs for different portions of the signal band. Each band is converted in parallel. A bank of FIR filters attenuates the out-

of-band noise for each band and achieves perfect reconstruction of the signal component. A design procedure is provided for the design of the filter bank with reduced computational complexity. The use of *complex* NTFs in the multi-band $\Delta\Sigma$ architecture is also proposed. The peformance of real and complex NTFs is compared. Performance evaluations are made for ideal systems as well as systems suffering from circuit implementation imperfections such as finite opamp gain and mismatched capacitor ratios.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Although real world signals are analog, it is often desirable to convert them into the digital domain using an analog to digital converter (ADC). Many reasons, such as virtually error free transmission and storage of the signal, motivate this conversion. For example, the digital representation of an audio signal allows CD players to achieve virtually error free storage using optical disks [1]. Intricate processing of the signal may also necessitate analog to digital (A/D) conversion since such processing is only feasible in the digital domain using either conventional digital computers or special purpose digital signal processors (DSPs). Also, in the biomedical field, signal processing such as the ultrasound imaging of human body organs based on analog signals acquired from sensors, can be performed with reliable accuracy only in the digital domain.

Analog to digital conversion of a signal is traditionally described in terms of two separate operations: uniform sampling in time and quantization in amplitude. Conventional A/D conversion involves sampling a signal just above the Nyquist rate and then independently quantizing each sample in amplitude. Such converters are known as Nyquist rate pulse code modulation (PCM) converters since their sampling frequency is near the Nyquist rate and their quantized amplitudes can be represented as digital codewords in a PCM format. The resolution of PCM quantization can be improved by *oversampling* the signal which uses a sampling frequency significantly higher than the Nyquist rate. *Noise shaping* in addition

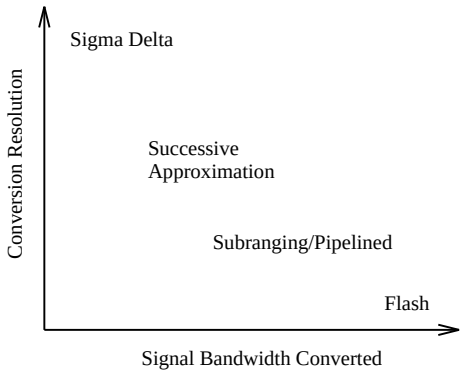


Figure 1.1: Bandwidth resolution tradeoffs for A/D converters

to oversampling has been used to perform high resolution A/D conversion using *only a 1 bit* internal A/D converter [2], [3], [4]. Such converters achieve their high resolution by oversampling, as well as processing the signal and quantization noise with different transfer functions. The signal transfer function (STF) is typically a delay over the signal band while the noise transfer function (NTF) is designed to shape the noise in such a way that it is attenuated in the signal band. Such converters are also widely known as sigma-delta ($\Sigma\Delta$) or delta-sigma ($\Delta\Sigma$) modulation converters. These, and other converters, provide tradeoffs among signal bandwidth converted, output resolution, and the complexity of the analog and digital hardware. Qualitative bandwidth and resolution tradeoffs of some of these A/D techniques as well as delta-sigma conversion is shown in Fig 1.1. As is evident from Fig 1.1, delta-sigma A/D converters attain the highest resolution of all the known conversion techniques but attain the lowest signal bandwidth. Consequently, delta-sigma techniques are often used in speech applications where the signal bandwidth is only 4 kHz and where up to 14 bits of resolution may be needed. Similarly, delta-sigma ADCs are popular for digital audio applications where the signal bandwidth is 20-24 kHz and where high fidelity (Hi-Fi) audio requires 16-18 bits of resolution. Flash converters on the other hand may be used for broadcast video applications where the signal band is about 5 MHz but the resolution required is only about 8 bits.

Recent research has focused on developing $\Delta\Sigma$ converter architectures which offer the potential of being able to push $\Delta\Sigma$ converter performance towards the upper middle if not the upper right hand corner of Fig 1.1 – i.e. to provide more resolution at higher

bandwidths. This thesis provides contributions towards this problem.

1.2 Background

Many engineering problems are solved by dividing or slicing the task into smaller sub-problems in various domains and then combining the results of the sub-problems to obtain an overall result. This is true for operations ranging from efficient sorting to efficient computation of the DFT. This thesis takes the view that this is no less true for analog to digital (A/D) conversion. The thesis examines A/D conversion based on several different slicing approaches.

For Nyquist rate PCM conversion, the quantization of each sample in amplitude can be performed by "slicing" the amplitude such that lower resolution converters can be used to quantize various "sliced" amplitudes. The results of the low resolution converters can then be combined to obtain a higher resolution overall output. Nyquist rate methods such as successive approximation, algorithmic, folding, and pipelined methods effectively take this approach of slicing in the amplitude domain.

The resolution achievable by a $\Delta\Sigma$ converter is a strong function of the oversampling ratio which is the ratio of the sampling frequency to the signal bandwidth. For a given signal bandwidth, the resolution obtained by $\Delta\Sigma$ converters can be improved by increasing the sampling frequency and hence the oversampling ratio. Increasing the sampling frequency is limited by the maximum circuit speed of the integrated circuit (IC) technology and the power dissipation. Many existing $\Delta\Sigma$ converters rely on architectural innovation rather than just increasing the oversampling ratio to improve the resolution. Such architectures involve using multi-bit internal quantizers, higher order single stage modulator structures as well as higher order cascaded structures. Improvements obtained using such architectures come at a price and have their limitations. For example, stability is a design issue for single stage higher order structures using an 1 bit internal quantizer and matching between successive stages becomes a design issue for higher order cascaded architectures. Consequently, there has been a need to research architectures which can be used as alternatives to or in conjunction with existing architectures.

This thesis researches both oversampled PCM and delta-sigma A/D converters which use architectural parallelism to employ slicing in the frequency or time domains. Frequency slicing involves quantizing different portions of the signal frequency spectrum using several quantizers in parallel and then combining the results of the quantizers to form an overall result. Time slicing involves quantizing various groups of time domain signal samples with different quantizers in parallel and then combining the results of the quantizers to form an overall output.

1.3 Contributions

The thesis makes contributions in the investigation of parallel frequency and time sliced A/D converter architectures by,

- Examining A/D conversion from the general context of time and frequency slicing and comparing the relative performance of time and frequency sliced PCM and $\Delta\Sigma$ systems. Frequency and time sliced systems for PCM conversion have existed for some time. However, their quantization noise performance have not been compared directly. Although it is known that the representation of a signal in the frequency and time domains are completely equivalent, the thesis analyzes the performance of frequency and time sliced A/D conversion systems and shows that the performance is not equivalent.
- Proposing the use of frequency slicing for delta-sigma A/D conversion.
- Proposing a reduced computational complexity digital FIR filter bank architecture to reconstruct the signal in a frequency sliced or mutli-band delta-sigma modulator system.
- Extending the use of time slicing for low-pass delta-sigma modulators to band-pass modulators.
- Proposing the use of complex noise transfer functions in delta-sigma modulators.
- Quantifying and comparing the performance of complex noise transfer functions with that of real noise transfer functions. The performance comparison is made for ideal

delta-sigma systems as well as for such systems suffering from implementation imperfections.

1.4 Thesis Organization

The thesis is organized as follows. A review of Nyquist rate PCM A/D conversion and oversampled PCM conversion are provided in Chapter 2. Chapter 3 reviews conventional delta-sigma conversion architectures. This thesis does not focus on aspects of the basic delta-sigma converter architecture and behavior such as robustness to circuit imperfections, non-linear behavior, stability, and decimation. However, for the sake of tutorial value, such issues are qualitatively discussed in Chapter 3 for some of the architectures which are reviewed. The concepts of slicing in general and amplitude slicing in particular are introduced in Chapter 4 using examples of Nyquist rate converters. Chapter 5 discusses oversampled PCM conversion from the point of view of frequency and time slicing. The performance of these systems is analyzed and it is shown that the performance is not equivalent. Simulation results are used to confirm the analysis. In Chapter 6, the use of complex noise transfer functions (NTFs) in band-pass $\Delta\Sigma$ modulators is introduced. The performance of the complex NTF is compared with that of real NTFs. Chapter 7 examines the practical performance of delta-sigma modulator architectures employing both complex and real transfer functions when implementation imperfections are present. Behavioral simulations are used to quantify the performance comparison. Time sliced delta-sigma conversion is examined in Chapter 8. The thesis extends the use of time slicing used in low-pass delta-sigma modulators to band-pass modulators. Simulation results are presented to validate this extension. In Chapter 9, the thesis proposes frequency sliced or multi-band delta-sigma conversion. The performance of real and complex NTFs in frequency sliced systems is shown. The thesis presents a method for the design of the digital filter bank in a frequency sliced $\Delta\Sigma$ system. Finally, the practical performance of the frequency sliced system is discussed. Chapter 10 summarizes some conclusions and discusses possibilities for future research.

Chapter 2

Review of PCM Conversion

2.1 Introduction

In this chapter, PCM analog to digital (A/D) conversion is reviewed. Analog to digital conversion of a signal is traditionally described in terms of two separate operations: uniform sampling in time and quantization in amplitude. The chapter starts with a basic discussion of sampling and quantization. It then describes how the performance of PCM A/D converters is most commonly modeled. This is followed by a discussion of the limitations faced by these converters. Finally, the generic architecture and performance of oversampled PCM converters are described.

2.2 Nyquist Rate PCM Conversion

2.2.1 Sampling and Quantization

Sampling

In the sampling process, a continuous time signal is sampled at uniformly spaced time intervals, T_s. The samples $x[n]$ of the continuous time signal $x(t)$ can be represented as $x[n] = x(nT_s)$. The effect, in the frequency domain, of the sampling process is to create periodically repeated versions of the signal spectrum at multiples of the sampling frequency $f_s = \frac{1}{T_s}$ [5, pp. 80-87]. This relationship is written in (2.1) where $X_s(f)$ represents the

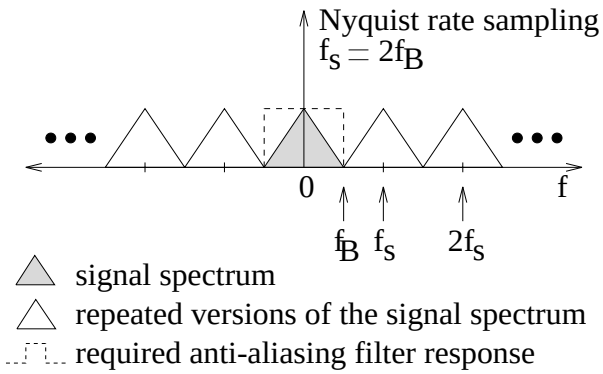


Figure 2.1: Nyquist rate sampling showing the original band limited signal spectrum, periodically repeated versions of the signal spectrum due to sampling, and the anti-aliasing filter response needed to band limit the signal.

spectrum of the sampled signal and $X(f)$ is the spectrum of the original continuous time signal.

$$X_s(f) = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} X(f - kf_s) \tag{2.1}$$

The sampling process is shown graphically in Fig 2.1 for the case where $f_s = 2f_B$ and f_B is the bandwidth of the signal. In general, the signal can be reconstructed back to continuous time if the repeated versions of the signal spectrum do not overlap. Thus, the signal must be band limited to half the sampling rate i.e. a signal with bandwidth f_B must be sampled at a rate greater than twice the bandwidth ($f_s \geq 2f_B$). Interference between the repeated versions of the signal spectrum is known as aliasing and it prevents reconstruction of the signal. Even if a signal is nominally band limited to $\frac{f_s}{2}$, an anti-aliasing filter is often used to ensure that the signal is indeed band limited. For example, speech has a nominal bandwidth of 4 kHz and so in principle can be sampled at 8 kHz. However, there is some residual signal energy above 4 kHz which results in aliasing if a 8 kHz sampling rate is used. The anti-aliasing filter is a continuous time analog filter preceding the sampler. The case where $f_s = 2f_B$ is known as Nyquist rate sampling and clearly the anti-aliasing filter here must have a very sharp cutoff at frequency $f_B = \frac{f_s}{2}$ as shown in Fig 2.1. Later we will discuss how the sharp cutoff requirement on this filter can be relaxed. Note that the discretization or quantization in time as a result of the sampling is an invertible operation since no signal information is lost and the original continuous time signal can be perfectly reconstructed [5, pp. 80-91]. We should finally note that although Fig 2.1 shows the sampling process for the

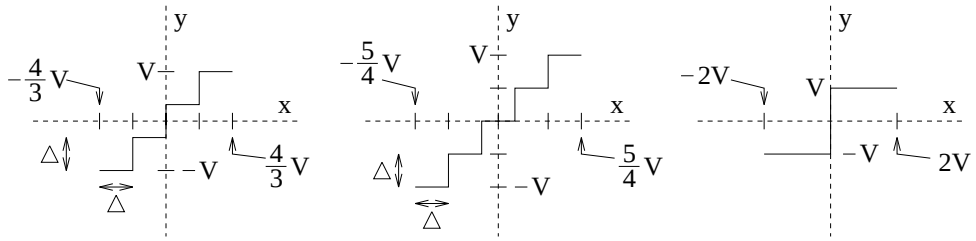


Figure 2.2: Transfer characteristics of typical A/D converters (ADCs) or quantizers

case where the signal is a baseband signal i.e. the spectrum has a bandwidth centered at DC frequency, equation (2.1) still describes the sampled spectrum even if the signal spectrum is centered at some higher frequency f_c. In this case, for a signal bandwidth f_B, the signal spectrum occupies the region, $[f_c - \frac{f_B}{2}, f_c + \frac{f_B}{2}]$, and it will still be possible to avoid aliasing and reconstruct the signal provided that $f_s \geq 2f_B$.

Quantization

Once sampled, the signal samples must also be quantized in amplitude to a finite set of output values. Typical transfer characteristics of quantizers or A/D converters with an input signal sample $x[n]$ and an output $y[n]$ are shown in Fig 2.2. Note that quantization is a non-invertible process since an infinite number of input amplitude values are mapped to a finite number of output amplitude values. The quantized output amplitudes are usually represented by a digital code word composed of a finite number of bits. For example, for the 1 bit A/D converter of Fig 2.2(c), the output levels V and $-V$ can be mapped to digital codes "1" and "0". The digital code words are often said to be in pulse code modulation (PCM) format. Another way of looking at this would be to plot the digital code words instead of the quantized amplitude values for y in Fig 2.2. The quantized output amplitude values can also be considered the output of an ideal digital to analog converter (DAC) whose inputs are the corresponding digital code words. An ADC or quantizer with Q output levels is said to have N bits of resolution where $N = log_2(Q)$. As should be clear from Fig 2.2, for an ADC with Q quantization levels, only input values separated by at least $\Delta = \frac{2V}{Q-1}$ can be distinguished or resolved to different output levels. N digital bits are needed to encode

the Q codewords corresponding with each output level. The difference between the binary digital codes for two adjacent output levels is one least significant bit (LSB) of the overall N bit codeword. Consequently, a Δ difference in input amplitude corresponds to a one LSB difference in the digital output code word.

We now point out some general properties of the quantization transfer curves shown in Fig 2.2(a) and (b) for a four level (two bit) "midriser" and a five level (roughly 2 bit) "midtread" ADC. Unlike the midtread ADC, the midriser ADC does not contain a zero output level for a zero input value – effectively creating a DC offset which may be undesirable in some applications. Note that all the transfer characteristics shown in Fig 2.2 are symmetric. The midriser needs to have an even number of output levels to produce a completely symmetric transfer curve whereas the midtread needs an odd number of output levels. The fact that the midriser ADC has a symmetric characteristic with an even number of levels is an advantage because the number of output levels, Q, can be made a power of two and encoded with exactly $N = \log_2(Q)$ bits. However, the number of output levels, Q, for a symmetric midtread ADC must be odd and so can not be made a power of 2 and encoded as efficiently. The number of bits needed will be $N = \log_2(Q-1)+1$ where $Q-1$ is chosen a power of 2. If the number of levels for the midtread ADC is forced to be a power of 2 by using only $Q-1$ levels, it will no longer have a symmetric transfer characteristic and will distort large amplitude symmetric input signals (e.g. a sinusoid). This distortion, of course, may be negligible when the number of output levels is very large. These issues may play a role in choosing whether a midriser or midtread quantizer transfer characteristic should be used. For the special case of a 2 level quantizer, a midtread characteristic will not be able to represent both positive and negative output levels and so will severely distort a signal containing samples of both polarities. For this 2 level case, a midriser characteristic, shown in Fig 2.2(c) will almost always be used.

2.2.2 Performance Modeling

Having looked at the sampling and quantization processes, let us now examine the A/D converter and try to characterize its performance. The diagrams in Fig 2.2 show the transfer characteristic for typical quantizers with input x and output y. Let the maximum and

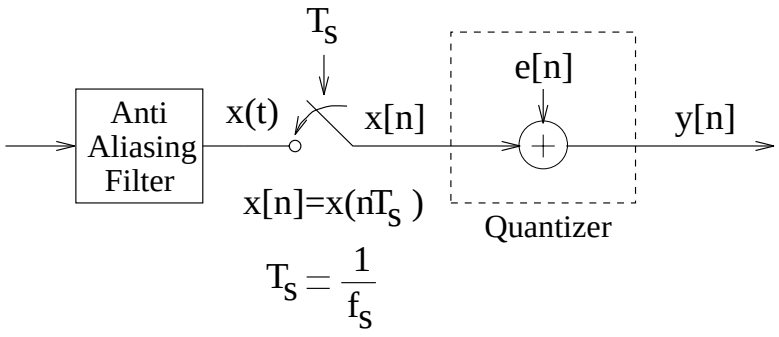


Figure 2.3: Block diagram and model of a conventional A/D converter (ADC) system

minimum quantized output values always be V and $-V$. The least significant bit (LSB) of an ADC with Q quantization levels is then equivalent to $\Delta = \frac{2V}{Q-1}$. For both the midriser and midtread type of ADCs of Fig 2.2, the magnitude of the quantization error ($e = y - x$) between the output and input does not exceed half a LSB i.e. $|e| \leq \frac{\Delta}{2}$ provided that $|x| \leq V + \frac{\Delta}{2}$. Under these circumstances, the quantizer or ADC is said to be not overloaded. For $|x| > V + \frac{\Delta}{2}$ (hence, $|e| > \frac{\Delta}{2}$) the ADC is said to be overloaded.

The quantizer embedded in any ADC is a non-linear system which makes its analysis difficult. To make the analysis tractable, the quantizer is often linearized and modeled by a noise source $e[n]$ being added to the signal $x[n]$ to produce the quantized output signal $y[n]$:

$$y[n] = x[n] + e[n] \tag{2.2}$$

A block diagram of an A/D system showing the sampling process and the quantizer model is shown in Fig 2.3. To further simplify the analysis of the noise from the quantizer, the following assumptions about the noise process and its statistics are traditionally made [5, p. 120].

1. The error sequence $e[n]$ is a sample sequence of a stationary random process.

2. $e[n]$ is uncorrelated with the sequence $x[n]$.

3. The probability density function of the error process is uniform over the range of quantization error i.e. over $\pm\frac{\Delta}{2}$.

4. The random variables of the error process are uncorrelated i.e. the error is a white noise process.

Under certain conditions, such as when the quantizer is not overloaded, N is large, and the successive signal values are not excessively correlated, these assumptions are reasonable [6]. Consider a N bit ADC with $Q = 2^N$ quantization levels i.e. with $\Delta = \frac{2V}{Q-1} = \frac{2V}{2^N-1}$. For a zero mean $e[n]$, its variance σ_e^2 or power is

$$\sigma_e^2 = \frac{\Delta^2}{12} = \left(\frac{2V}{2^N - 1}\right)^2 /12 \simeq \left(\frac{2V}{2^N}\right)^2 /12 = \frac{V^2}{3\,(2^{2N})} \tag{2.3}$$

If the signal is treated as a zero mean random process and its power is σ_x^2 then the signal to quantization noise ratio in dB is

$$SNR = 10\,log\left(\frac{\sigma_x^2}{\sigma_e^2}\right) = 10\,log\frac{\sigma_x^2}{V^2} + 4.77 + 6.02N\ (dB) \tag{2.4}$$

Note that for each extra bit of resolution in the ADC i.e. for every increment in N, there is about a 6 dB improvement in the SNR. Thus, there is a direct relationship between the resolution of an ADC and its SNR and it is common to equate differences in SNR in dB to bits by dividing the dB value by 6. For example, if an ADC has a SNR which is 3 dB better than that of another ADC, the better ADC is said to have $\frac{1}{2}$ bit higher resolution. Also, note that for a given N, the SNR in dB is linearly related to the signal power, σ_x^2, in dB.

Let us now examine the dynamic range of the ADC which is a measure of the range of input amplitudes for which the ADC produces a positive SNR. For sinusoidal inputs, the dynamic range of the A/D converter is defined as the ratio of the signal power of a full scale sinusoid to the signal power of a small sinusoidal input which results in a SNR of 0 dB [7]. The signal power of a full scale sinusoid is $\frac{V^2}{2}$. A sinusoid with signal power $\sigma_x^2 = \sigma_e^2 = \frac{\Delta^2}{12}$ will result in the SNR being 1 or 0 dB. The dynamic range, by definition, is then $\left(\frac{V^2}{2}/\frac{\Delta^2}{12}\right)$ $\simeq \left(\frac{V^2}{2}/\frac{\left(\frac{2V}{2^N}\right)^2}{12}\right)$. This reduces to a dynamic range value given by:

$$D_R = 6.02N + 1.76\ (dB) \tag{2.5}$$

Note that the ratio of $\frac{V^2}{2}$ to $\frac{\Delta^2}{12}$ is just the peak SNR of the ADC for a sinusoidal input. Consequently, the dynamic range of the Nyquist rate ADC is the same as its peak SNR.

We will later see that delta-sigma converters do not necessarily have their peak SNR equal to their dynamic range. However, by using the dynamic range of a delta-sigma converter in (2.5) and calculating the corresponding N we will be able to determine the resolution of a Nyquist rate PCM converter which would be required to produce the same dynamic range.

2.2.3 Limitations of Nyquist Rate ADCs

For Nyquist rate converters, each signal sample is quantized at the full precision or resolution of the converter. The resolution of such converters implemented on VLSI chips is limited by the technology in which these chips are fabricated. For example, some successive approximation A/D techniques rely on matching of two capacitors to perform a repeated division of a reference voltage by 2. If such a converter is to convert signal values to N bits of resolution, the required matching on the capacitor components needs to be at least one part in 2^N. Matching of components to greater than 10 bits (one part in 2^{10}) or equivalently to more than 0.1% is difficult in VLSI. High resolution Nyquist rate converters are extremely difficult to attain in current integrated circuit (IC) technology without the use of techniques such as laser trimming of components or calibration. Furthermore, if the signal is sampled too close to the Nyquist rate, the anti-aliasing filter must have a very sharp cutoff which is a non-trivial design requirement for analog filters.

2.3 Oversampled PCM Conversion

2.3.1 System Description

Oversampled PCM conversion is a technique which improves the resolution obtained from straightforward Nyquist rate PCM conversion. This is achieved by oversampling the signal, i.e. samples are acquired from the analog waveform significantly faster than the Nyquist rate. Each of these samples is quantized by a N bit ADC. Since quantization is described by (2.2), which is $y[n] = x[n] + e[n]$, the total amount of noise power injected into the sampled signal, $x[n]$, is σ_e^2 and is given by (2.3). Obviously, this is exactly the same noise power produced by a Nyquist rate converter but its frequency distribution relative to the signal is different because of the higher sampling rate. Assumption (4) in Sec 2.2.2 designated the

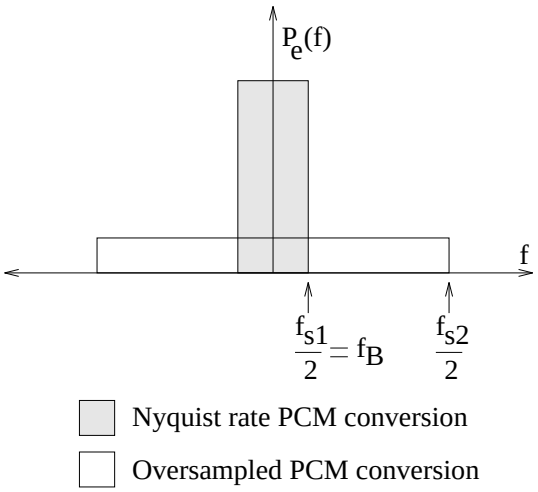


Figure 2.4: Quantization noise power spectral density for Nyquist rate PCM and oversampled PCM conversion

noise process as white which means that the noise power is uniformly distributed between $-\frac{f_s}{2}$ to $\frac{f_s}{2}$ where f_s is the sampling frequency. Fig 2.4 shows the power spectral density, $P_e(f)$, of the quantization noise for Nyquist rate sampling with rate f_{s1} and oversampling with rate f_{s2}. For Nyquist rate sampling where the signal band, $f_B = \frac{f_{s1}}{2}$, all the quantization noise power, represented by the area of the tall shaded rectangle, occurs across the signal bandwidth. In the oversampled case, the same noise power, represented by the area of the unshaded rectangle has been spread over a bandwidth equal to the sampling frequency f_{s2} which is much greater than the signal bandwidth f_B. Only a relatively small fraction of the total noise power falls in the band $[-f_B, f_B]$, and the noise power outside of the signal band can be greatly attenuated with a digital low-pass filter following the ADC. After the low-pass filtering is performed, the signal can be downsampled to the Nyquist rate without affecting the signal to noise ratio. The collective operation of low-pass filtering and downsampling is known as decimation. A block diagram of an oversampled PCM system showing the sampling, the ADC model, and the decimator is presented in Fig 2.5.

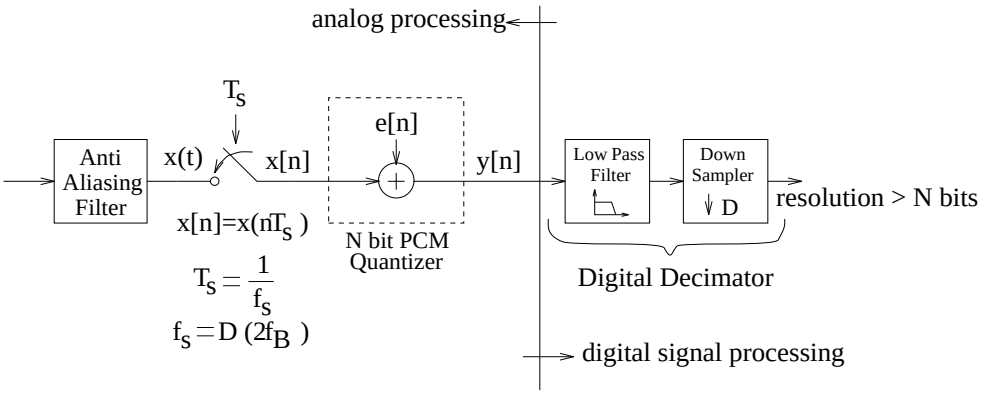


Figure 2.5: Oversampled PCM conversion system

2.3.2 Performance Modeling

By taking the Z transform of (2.2), the Z domain relationship between the input and output of an oversampled PCM converter is obtained as,

$$Y(z) = X(z) + E(z) \tag{2.6}$$

where Y, X and E are the Z transforms of the output, input signal and the quantization error process, respectively. Based on our two-input linear system model for the quantizer, (2.6) states that in the Z domain, the output is the input plus the quantization error or noise. We can observe that X and E both experience a unity transfer function. A more general way of writing (2.6), where X and E do not necessarily experience a unity transfer function, is

$$Y(z) = X(z)H_x(z) + E(z)H_e(z) \tag{2.7}$$

The output is now the input signal modulated by a signal transfer function (STF), denoted by $H_x(z)$, plus the quantization noise modulated by a noise transfer function (NTF), denoted by $H_e(z)$. To evaluate the performance of such a converter, we need to find the total signal and noise power at the output of the converter. To do this, we need to evaluate the power spectral densities $P_{xy}(f)$ and $P_{ey}(f)$ of the signal and noise at the output of the converter, based on the power spectral densities, $P_x(f)$ and $P_e(f)$ of the signal and noise at the input of the converter. We can make use of the fact that if a stationary random process with power spectral density $P(f)$ is the input to a linear filter with transfer function $H(f)$, the

power spectral density of the output random process is $P(f)|H(f)|^2$. Consequently,

$$P_{xy}(f) = P_x(f)|H_x(f)|^2$$

$$P_{ey}(f) = P_e(f)|H_e(f)|^2$$

For the oversampled PCM converter, $|H_x(f)| = |H_e(f)| = 1$ and our white noise assumption for $e[n]$ states that $P_e(f) = \frac{\sigma_e^2}{f_s}$ which implies $P_{ey}(f) = \frac{\sigma_e^2}{f_s}$. Assuming an *ideal* low-pass filter with cutoff frequency f_B following the oversampled quantizer, the *in-band* noise power, σ_{ey}^2, at the output of the A/D is

$$\sigma_{ey}^2 = \int_{-f_B}^{f_B} P_{ey}(f)df = 2\int_0^{f_B} P_{ey}(f)df = \int_0^{f_B} \frac{2\sigma_e^2}{f_s}df = \sigma_e^2\left(\frac{2f_B}{f_s}\right)$$

Note that some of the noise power is now located outside of the signal band as a result of the oversampling and so the in-band power σ_{ey}^2 is less than what it would have been without any oversampling (σ_e^2). Since the signal power is assumed to occur over the signal band only, it is not modified in any way and the signal power at the output σ_{xy}^2 is the same as the input signal power σ_x^2. The maximum achievable SNR $\frac{\sigma_x^2}{\sigma_{ey}^2}$ in dB is then,

$$SNR = 10\ \log\left(\frac{\sigma_x^2}{\sigma_{ey}^2}\right) = 10\,log(\sigma_x^2) - 10\,log(\sigma_e^2) + 10\,log\left(\frac{f_s}{2f_B}\right)\ (dB) \tag{2.8}$$

For the case of Nyquist rate sampling where $f_s = 2f_B$, this formula reduces to (2.4) which is the SNR for the Nyquist rate PCM quantizer. Letting the oversampling ratio, $\frac{f_s}{2f_B} = 2^r$, we obtain,

$$SNR = 10\,log(\sigma_x^2) - 10\,log(\sigma_e^2) + 3.01r\ (dB) \tag{2.9}$$

For every doubling of the oversampling ratio i.e. for every increment in r, the SNR improves by about 3 dB or the resolution improves by $\frac{1}{2}$ bit.

Note that in this scheme, we are trading speed for resolution. The higher resolutions are obtained at the expense of requiring the internal PCM quantizer to quantize samples at the oversampled rate. Analog circuit complexity has also been traded for digital circuit complexity. The analog circuit complexity is simplified since we have said that the resolution of the internal N bit quantizer, an analog circuit, is lower than that of the overall conversion resolution. Another benefit, which is a direct consequence of the oversampling, is that the

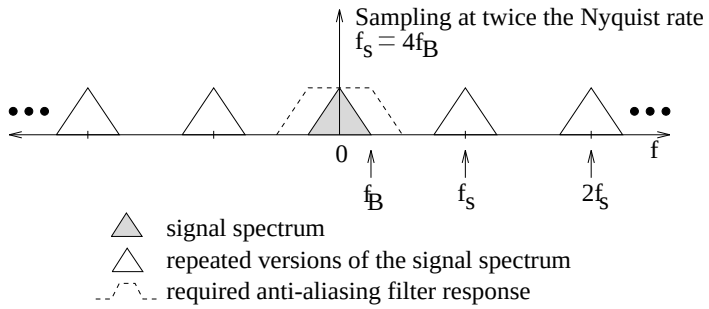


Figure 2.6: Sampling at twice the Nyquist rate showing the original band limited signal spectrum, repeated versions of the signal spectrum due to sampling, and an anti-aliasing filter which is sufficient to band limit the signal.

analog anti-aliasing filter does not need as sharp a cutoff. This can be seen from Fig 2.6 where a signal is sampled at four times the nominal signal bandwidth or at twice the Nyquist rate. In this case, the anti-aliasing filter can have a transition band between f_B and $\frac{f_s}{2}$ as long as it provides very good attenuation beyond $\frac{f_s}{2}$. However, a price is paid in the digital domain since the digital filter must attenuate the remaining quantization noise power (beyond f_B) as much as possible. In the process of filtering out-of-band quantization noise, any other noise which existed in the transition band of the anti-aliasing filter prior to sampling will be attenuated further. The closer the low-pass filter approximates an ideal low-pass filter, the more resources it will need.

Finally, note that for high resolution conversion, one needs, $f_s \gg f_B$, and the signal bandwidth must be small so that f_s does not exceed the maximum circuit speed attainable in the given technology. Let us use (2.8) for a simple calculation. Suppose we apply a full scale sinusoid with amplitude $V = 1$ corresponding with signal power $\frac{V^2}{2} = 0.5$, as the input to an oversampled PCM ADC with a 20 kHz audio range signal band. Let the final desired resolution be 16 bits (for CD quality audio) which corresponds to a 98 dB SNR according to (2.4). Now, if we use an 8 bit A/D in an oversampled PCM scheme, i.e. if we use $N = 8$ in the expression for σ_e^2 in (2.3), we can calculate the f_s required for SNR=98 dB and $f_B = 20$ kHz using (2.8). The needed f_s is 2.64 GHz ! 8 bit ADCs implemented in current CMOS technology certainly can not operate at such a high speed. Suppose a 12 bit internal ADC is used instead. In this case, the required f_s is about 10 MHz, an operating speed which is not extremely high, per se, but which is still not trivial for a 12 bit ADC to attain. We will

later see how $\Delta\Sigma$ modulation A/D conversion allows the use of internal ADCs with as low as **1 bit** (i.e. $N = 1$) of resolution to achieve an overall resolution of 16 bits for the 20 kHz audio bandwidth.

Chapter 3

Review of Delta-Sigma A/D Conversion

3.1 Introduction

In this chapter, delta-sigma A/D converters, which employ the idea of noise shaping, are discussed. Various $\Delta\Sigma$ A/D architectures are reviewed. The chapter begins with a discussion of the idea of noise shaping and then describes the simplest $\Delta\Sigma$ system – a first order $\Delta\Sigma$ modulator. The system is analyzed in the frequency domain and a qualitative description of the time domain behavior is also provided. Various other architectures reviewed are second and higher order modulators, multi-bit modulators, as well as multi-stage (cascaded) modulators. The use of delta-sigma modulation for the conversion of band-pass signals is also reviewed. The performance of these $\Delta\Sigma$ systems is characterized. Finally, decimation for delta-sigma modulators is briefly described.

3.2 Noise Shaping A/D Converters

A general way of writing the Z domain output of an A/D converter was given in (2.7) as $Y(z) = X(z)H_x(z) + E(z)H_e(z)$ where H_x is the signal transfer function (STF) and H_e is the noise transfer function (NTF). For oversampled PCM conversion we saw that $H_x(z) = H_e(z) = 1$. This need not be the case and, in fact, oversampled A/D converters can

be designed to incorporate *noise shaping* which means that H_e is designed to be different from H_x such that H_x usually leaves the signal undisturbed but H_e shapes the noise to allow a high resolution output [2], [3], [4]. Although the term sigma-delta ($\Sigma\Delta$) is also almost synonymous with noise shaping ADCs, we will use the term $\Delta\Sigma$ which was used by some of the earliest researchers in the field [8]. As noted, oversampling reduces the quantization noise power in the signal band by spreading a fixed quantization noise power over a bandwidth much larger than the signal band. Noise shaping or modulation further attenuates this noise in the signal band and has the side effect of amplifying it outside the signal band. Consequently, this process of noise shaping by the delta-sigma modulator can be viewed as *pushing* quantization noise power from the signal band to other frequencies. The modulator's digital output can then be low-pass filtered to attenuate the out-of-band quantization noise and finally can be downsampled to the Nyquist rate. The price of attaining high resolution is again a penalty in speed, as the hardware has to operate at the oversampled rate, and an increased complexity of the digital hardware. For high resolution conversion, the sampling frequency f_s must still be much greater than the signal bandwidth f_B, but, not as great as needed by oversampled PCM conversion.

3.3 First Order $\Delta\Sigma$ Modulation

3.3.1 Operation and Performance Modeling

A block diagram of a first order delta-sigma modulator A/D system is shown in Fig 3.1. The system consists of an analog $\Delta\Sigma$ modulator followed by a digital decimator. The modulator consists of an integrator, an internal A/D converter or quantizer, and a D/A converter or DAC used in the feedback path. The signal that is quantized is not the input $x[n]$ but a filtered version of the difference between the input and an analog representation, $y_a[n]$, of the quantized output $y[n]$. The filter, often called the feedforward loop filter is a discrete time integrator whose transfer function is $\frac{z^{-1}}{1-z^{-1}}$. The integrator and the rest of the $\Delta\Sigma$ analog circuit are typically implemented in sampled data switched capacitor technology. Consequently, the sampling operation is not shown explicitly in Fig 3.1 or any other modulator architectures to be described in the rest of the paper. Continuous time

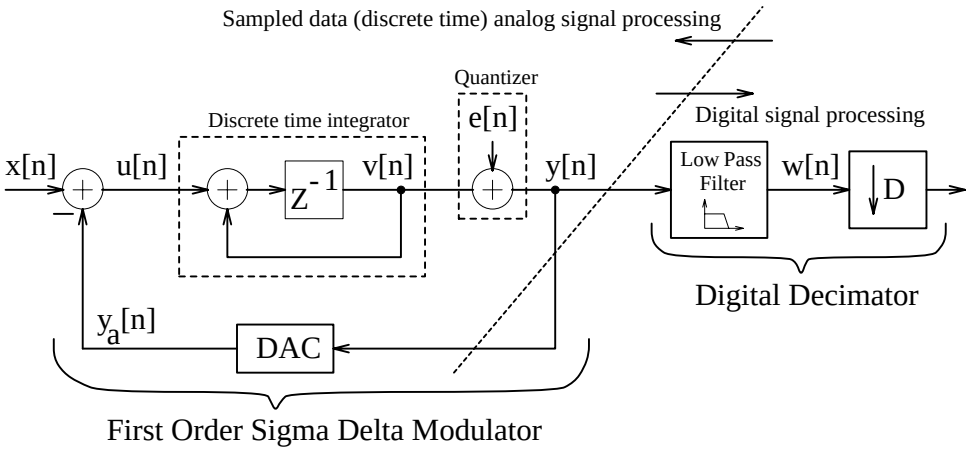


Figure 3.1: First order delta-sigma modulator A/D system

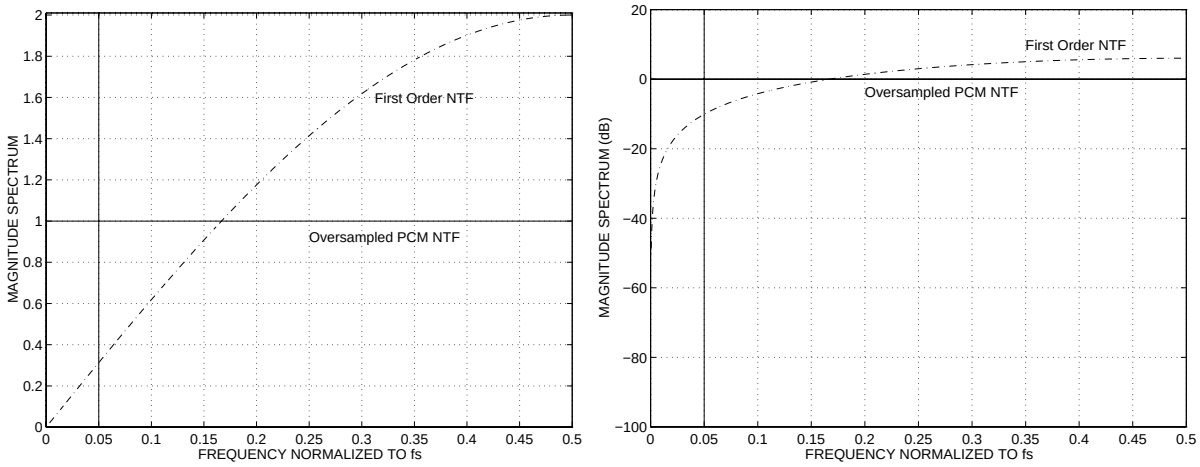
versions of the modulator have also been considered [9], but this aspect of the modulators will not be discussed here. The linearized model replaces the quantizer with a noise source, $e[n]$, as shown in Fig 3.1. If the DAC is ideal, it is replaced by a unity gain transfer function. The modulator output $Y(z)$ is then given by

$$Y(z) = X(z)z^{-1} + E(z)(1 - z^{-1}) \tag{3.1}$$

so that $H_x(z) = z^{-1}$ and $H_e(z) = (1 - z^{-1})$. The output is just a delayed version of the signal plus quantization noise that has been shaped by a first order Z domain differentiator or high-pass filter. The corresponding time domain version of the modulator output is

$$y[n] = x[n-1] + e[n] - e[n-1] \tag{3.2}$$

where the $e[n] - e[n-1]$ term is the first order difference of $e[n]$. The magnitude spectrum of a 1st order $\Delta\Sigma$ noise transfer function (NTF) is plotted in Fig 3.2(a) while Fig 3.2(b) shows the same plot in dB. The frequency axis has been normalized with respect to the sampling frequency, f_s. Since $H_e(z)$ contains a zero at $z = 1$ i.e. at DC frequency on the unit circle of the Z plane, note the zero gain or infinite attenuation provided by the NTF at DC frequency. Note the large attenuation at lower frequencies and relative amplification at higher frequencies. For comparison, the oversampled PCM NTF, which is unity gain, is shown in Fig 3.2(a). The vertical bar demarcates the extent of the signal band, f_B, where $f_B = 0.05 f_s$. Quantization noise to the left of the bar that contributes to the finite



(a) (b)

Figure 3.2: NTF for a 1st order delta-sigma modulator (a) Magnitude spectrum on linear scale. For comparison, the oversampled PCM NTF, which is a unity gain, is shown. (b) Magnitude spectrum in dB. The vertical bar demarcates the extent of the signal band, f_B, where $f_B = 0.05 f_s$ in this example.

resolution of the modulator is greatly attenuated while noise to the right of the bar is not attenuated as much or is actually amplified. However, noise to the right of the bar can mostly be removed with a digital low-pass filter.

Let us now say a few words about the DAC before going on to examine the performance of the converter. The DAC is required to be nearly as linear as the overall conversion resolution. This can be inferred from observing that any D/A non-linearity can be modeled as another error source which adds directly to the input. This error source benefits from the oversampling but unlike $e[n]$, which models the A/D quantization error, is not subject to the noise shaping. Since an 1 bit DAC is perfectly linear, it is common to use an 1 bit DAC and a corresponding 1 bit quantizer which is simply a comparator. Consequently, provided the sampling frequency is high enough, the ΔΣ A/D allows the use of an 1 bit quantizer to achieve high overall resolution. Using $H_x = z^{-1}$, $H_e = (1 - z^{-1})$, and the procedure used for the oversampling PCM A/D, the in-band noise power (i.e. the noise in the frequency range $[-f_B, f_B]$) at the output of a first order ΔΣ modulator is

$$\sigma_{ey}^2 = \sigma_e^2 \frac{\pi^2}{3} \left(\frac{2 f_B}{f_s} \right)^3 \tag{3.3}$$

The SNR in dB is then,

$$SNR = 10\,log(\sigma_x^2) - 10\,log(\sigma_e^2) - 10\,log\left(\frac{\pi^2}{3}\right) + 30\,log\left(\frac{f_s}{2f_B}\right)\ (dB) \tag{3.4}$$

Letting the oversampling ratio, $\frac{f_s}{2f_B} = 2^r$, we obtain

$$SNR = 10\,log(\sigma_x^2) - 10\,log(\sigma_e^2) - 10\,log\left(\frac{\pi^2}{3}\right) + 9.03r\ (dB) \tag{3.5}$$

For every doubling of the oversampling ratio i.e. for every increment in r, the SNR improves by 9 dB or the resolution improves by 1.5 bits. Let us revisit the example considered at the end of the discussion on oversampled PCM. It was desired to convert a 20 kHz audio band to CD quality resolution of 16 bits. Using (3.4), we can compute that the required f_s with an 1 bit internal ADC is 96.78 MHz. An 1 bit ADC or comparator can operate at this speed in current CMOS technology. However, it is not possible for the sampled data analog switched capacitor integrators to operate at such high speeds. Second order delta-sigma modulation which will be discussed later, will allow the use of an 1 bit quantizer allowing us to meet the 16 bit 20 kHz target with a much more reasonable f_s.

3.3.2 Qualitative Time Domain Behavior

The ΔΣ modulator can be thought of as a PCM converter with feedback, which attempts to force the output signal $y[n]$ to be equal to the input signal $x[n]$. Consider the case where an one bit A/D converter or comparator is used. The transfer characteristics of this 1 bit ADC with output levels V and $-V$ is shown in Fig 2.2(c). From now on let us assume that $V = 1$ such that the comparator digital output is 1 or -1 so that $y[n]$ and $y_a[n]$ can be used interchangeably. In the time domain, referring to Fig 3.1, we have,

$$v[n] = u[n-1] + v[n-1] \tag{3.6}$$

$$y[n] = \begin{cases} 1 & v[n] \geq 0 \\ -1 & v[n] < 0 \end{cases} \tag{3.7}$$

$$u[n] = x[n] - y[n] \tag{3.8}$$

The "error" between the modulator output and input is $u[n]$. Note that this is not the quantization error which is given by $e[n] = y[n] - v[n]$. Since $y[n]$ can take on values of 1

or -1 only, it can never equal the input unless the input happens to be one of these two values exactly. Consequently, except for the mentioned cases, there will always be an error $u[n] \neq 0$. Consider a DC input for $x[n]$. When $y[n] = 1$, $y[n]$ is greater than the input $x[n]$ and the error $u[n]$ is negative and so negative values are accumulated by the integrator to produce $v[n]$. After a number of clock cycles, enough negative values will have accumulated to cause the quantizer to produce $y[n] = -1$ thereby changing the sign of the error $u[n]$ to be positive. The error between the output and input has been reduced, in some sense, because the positive errors will now cancel the prior negative errors when averaged over a period of time. Now with $y[n] = -1$, the errors will be positive and positive values of the error will be accumulated again until the quantizer output changes, this time back to $y[n] = 1$. Over a period of time, the proportion (or density) of 1's and -1's will be related to the DC input value – the larger the input, the more 1's will be present in the output and vice versa for smaller inputs. For this reason, the output of a $\Delta\Sigma$ modulator using an 1 bit quantizer is often said to be in pulse density modulated (PDM) format. Let us now illustrate the time domain behavior using a few examples. Fig 3.3(a) shows a DC input $x[n] = 0.55 = \frac{11}{20}$ while Fig 3.3(b) shows the corresponding modulator output $y[n]$. Roughly, three fourth of the output values are 1's while the others are -1's. Fig 3.3(c) shows the error signal $u[n]$ while Fig 3.3(d) shows the accumulated error signal or integrator output $v[n]$ whose sign change forces the quantizer output to change. For comparison, for a DC input of $x[n] = 1$, all the modulator output values will be 1's while for a zero DC input, half the modulator output values will be 1's, half -1's and for a DC input of -1, all the values will be -1's.

By averaging the modulator output over a period of time we can approximate the input. This averaging operation represents the low-pass filter block in Fig 3.1 since averaging is a crude low-pass filtering operation. Using a better low-pass filter will result in the modulator output being a better approximation to the input for a given oversampling ratio.

Finally, let us look at some time domain examples of the modulator output for sinusoidal inputs. Fig 3.4(a)–Fig 3.4(d) show the modulator outputs, $y[n]$, for various sinusoidal inputs. As for the DC input case, the sinusoid amplitude information is encoded in the relative number of 1's vs -1's. The modulator output pulse pattern has periodic components and the fundamental period encodes the sinusoid frequency. This is particularly clear in

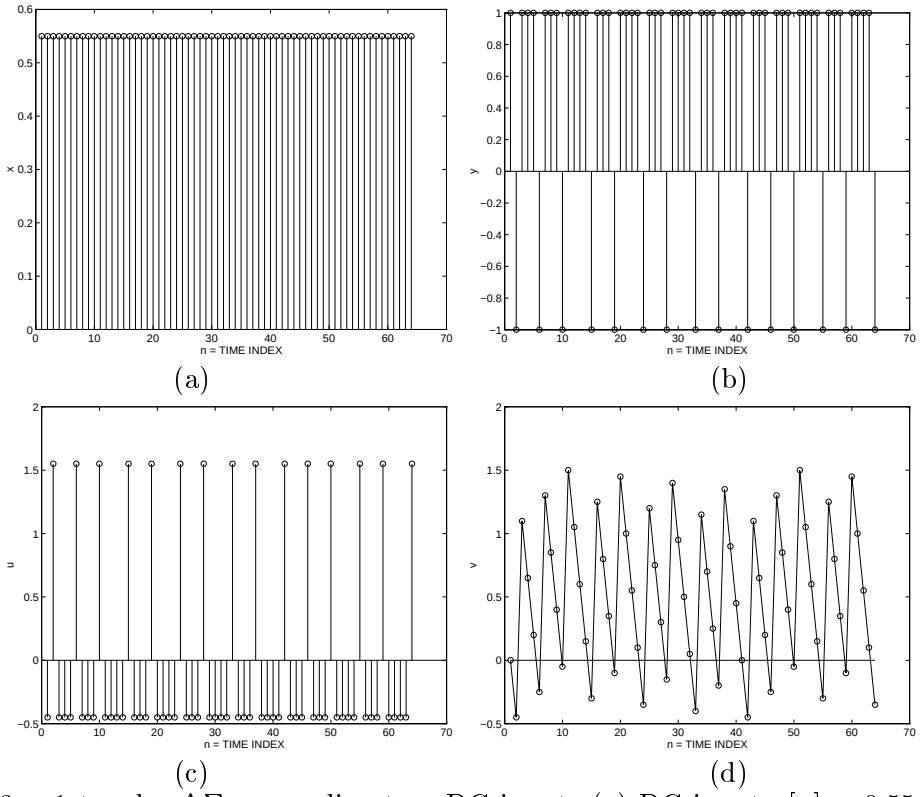


Figure 3.3: 1st order ΔΣ responding to a DC input: (a) DC input $x[n] = 0.55 = \frac{11}{20}$ (b) modulator output $y[n]$ (c) "error" signal $u[n]$ (d) integrator output $v[n]$

Figs 3.4(a) and (c).

3.3.3 Implementation Imperfections

The results presented so far have not taken into account imperfections in the analog hardware. Let us now discuss the consequences of imperfections in some of the main circuit parameters.

The integrator in the modulator may have a gain of g instead of unity and may be leaky. For an input $u[n]$, an integrator with gain g and leakage factor α has an output $v[n] = gu[n-1] + \alpha v[n-1]$ instead of $v[n] = u[n-1] + v[n-1]$ and the integrator transfer function is $\frac{gz^{-1}}{1-\alpha z^{-1}}$ instead of $\frac{z^{-1}}{1-z^{-1}}$. The D/A gain may also not be perfectly unity and assuming a gain of d, the STF and NTF are $H_x = \frac{gz^{-1}}{1+(gd-\alpha)z^{-1}}$ and $H_e = \frac{(1-\alpha z^{-1})}{1+(gd-\alpha)z^{-1}}$.

The original NTF $(1-z^{-1})$ which had a Z domain zero at $z = 1$ (on the unit circle and at DC) now has a zero which is still at DC but is moved inside the unit circle. This degrades

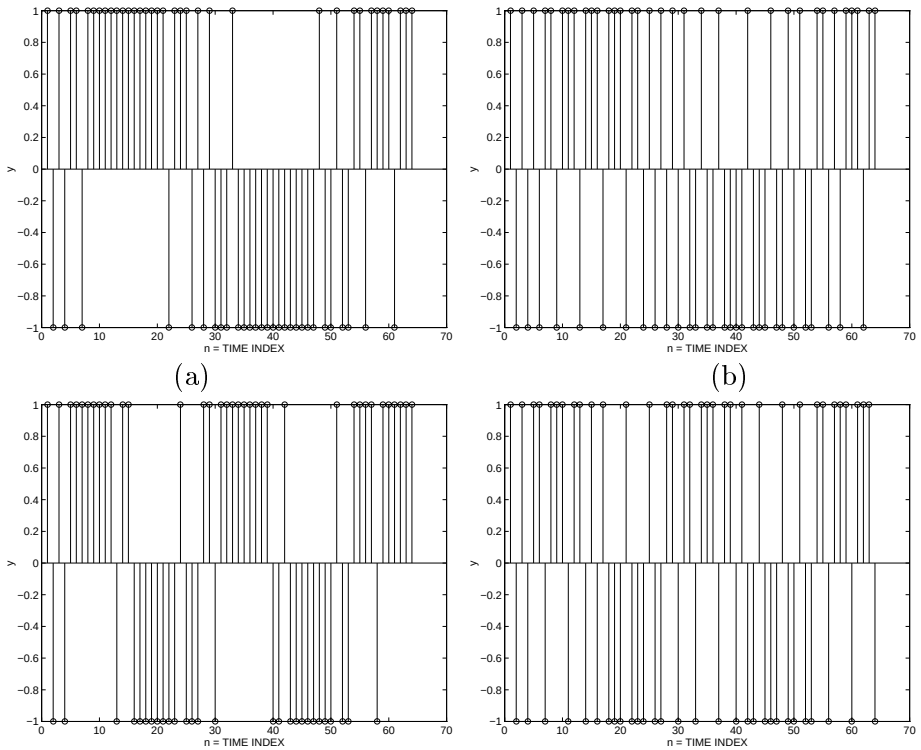


Figure 3.4: Modulator output $y[n]$ of a 1st order $\Delta\Sigma$ responding to various sinusoidal inputs with sampling frequency of about 1 MHz: (a) amplitude=0.95, frequency=20 kHz (b) amplitude=0.5, frequency=20 kHz (c) amplitude=0.95, frequency=40 kHz (d) amplitude=0.5, frequency=40 kHz

the NTF noise attenuation in the signal band and can thus affect the noise performance significantly. The term "leaky" comes from the fact that there is charge leakage in the switched capacitor implementation of the integrator circuit and only a portion (α) of the charge from the input capacitor is transferred to the integrating capacitor. The leakage factor α is related to the open loop gain, A, of the operational amplifier (opamp) used to implement the switched capacitor integrator such that $1-\alpha \simeq \frac{1}{A}$. Ignoring the denominator (which will add a slight ripple to the numerator) of the degraded H_e i.e. by considering the degraded NTF to be $(1-\alpha z^{-1})$, the in-band quantization noise power with integrator leakage is

$$\sigma_{ey}^2 = \sigma_e^2 \left(\frac{2f_B}{f_s}\right) \frac{1}{A^2} + \alpha\sigma_e^2 \frac{\pi^2}{3} \left(\frac{2f_B}{f_s}\right)^3 \tag{3.9}$$

instead of (3.3). Note that the noise power now contains a term which is inversely proportional to the oversampling ratio, $\frac{f_s}{2f_B}$, as well as a term inversely proportional to the

oversampling ratio cubed. However, the first term is divided by A^2 and it has been found [2] that if the opamp open loop gain, A, exceeds the oversampling ratio, $\frac{f_s}{2f_B}$, leakage causes no significant degradation of the SNR [2]. Consequently, the circuit constraint required to implement good integrators is not that difficult to meet unless the oversampling ratio is extremely high.

From the linearized analysis, the STF and NTF pole is stable only for $0 < gd < 2$. Consequently, there is a relatively wide margin over which the two gains may vary from this point of view. It has been reported [2] that variations of g by up to 10% from unity does not degrade the SNR significantly. The gain is implemented in practice as a ratio of two capacitors and so the corresponding precision on the capacitor matching is minimal – one part in 10 or three and a quarter bits. Now consider an imperfect DAC gain d that is slightly different from unity. This can be modeled as a gain of $\frac{1}{d}$ at the input of the modulator. To see this, consider that a gain of $\frac{1}{d}$ is inserted at the input of the integrator. This gain can be moved past the summing node at the modulator input in Fig 3.1. The result is that the DAC gain d is cancelled by the gain $\frac{1}{d}$ but the input now experiences a gain of $\frac{1}{d}$ before the modulator. Consequently, the STF experiences a slight gain change but there is no great impact on the modulator SNR.

Finally, consider imperfections in the quantizer. Any non-linearity in the quantizer can be modeled as another noise source which is added to $e[n]$, the quantization error. However, the noise from this extra source is subject to noise shaping by the modulator and so its affect on SNR degradation is not significant. If the 1 bit quantizer or comparator has a non-zero threshold, v_{th}, its output is given by:

$$y[n] = \begin{cases} 1 & v[n] - v_{th} \geq 0 \\ -1 & v[n] - v_{th} < 0 \end{cases}$$

This is equivalent to an offset at the input of the comparator i.e. at the output of the integrator. However, an offset v_{th} at the output $v[n]$ of the integrator corresponds to a discrete time impulse, $\delta[n]$, at its input $u[n] = x[n] - y[n]$ which amounts to one incorrect output $y[n]$ for a given $x[n]$. One such incorrect value will have a negligible impact on the overall performance of the modulator. The offset can also be modeled as an error source, with mean v_{th} and zero variance, that adds to $e[n]$. If for any reason, the offset is

input dependent or changes with time, this new error source will have non-zero variance. However, it will be subject to the noise shaping property of the modulator just as $e[n]$, and its presence will not degrade the performance of the modulator significantly.

3.3.4 Non-linear Behavior

A delta-sigma modulator is a non-linear system incorporating feedback. Not surprisingly, the modulator may display limit cycle oscillations which result in the presence of periodic (tone) components being present in the output and in the quantization error. This phenomenon is analogous to limit cycles which occur in digital IIR filters operating with finite precision arithmetic, because, like a $\Delta\Sigma$ modulator, such a filter is a non-linear system which employs feedback. The fact that the quantizer error spectrum is not white is again not surprising as the conditions for the white noise assumption are not perfectly satisfied – the quantizer has only two output levels, and due to the oversampling, successive quantizer input samples may be correlated.

Let us illustrate the existence of limit cycles in the modulator as has been done in [10] for the simple case of a DC input, $x[n] = x$. For a limit cycle of period T, $v[n]$ should be periodic with period T i.e. $v[n] = v[n+T]$. This clearly implies that $y[n] = y[n+T]$. For the DC input, the input to the integrator, $u[n] = x - y[n]$, will likewise be periodic with period T. Thus, the modulator behavior can be represented by T equations. Combining (3.6) and (3.8) we obtain $v[n] - v[n-1] = x[n-1] - y[n-1]$ which for a DC input becomes $v[n] - v[n-1] = x - y[n-1]$. Writing this equation for T different time instances starting (arbitrarily) at $n = 1$, we obtain

$$\begin{aligned}
v[1] - v[0] &= x[0] - y[0] \\
v[2] - v[1] &= x[1] - y[1] \\
&\vdots \\
v[T-1] - v[T-2] &= x[T-2] - y[T-2] \\
v[T] - v[T-1] &= x[T-1] - y[T-1]
\end{aligned}$$

Adding up all these equations we obtain,

$$v[T] - v[0] = \sum_{l=0}^{T-1} x - \sum_{l=0}^{T-1} y[l] \tag{3.10}$$

However $v[T] = v[0]$ by assumption and consequently,

$$x = \frac{1}{T}\sum_{l=0}^{T-1} y[l] = \frac{(P-M)}{T}V \tag{3.11}$$

where P is the number of positive quantizer outputs over T output samples and M is the number of negative quantizer outputs over T output samples. Since $(P-M)$ is an integer as is T by assumption, we have $x = \frac{b}{a}V$ with b, a integers. Thus, a necessary condition for the output y to consist of a limit cycle with period T is that the DC input x is a rational multiple of V. The limit cycle with period T will manifest itself in the output spectrum as tones present at frequency $\frac{f_s}{T}$ and its harmonics. The period is $T = 2a$ if b or a is even but is $T = a$ if both a and b are odd [12]. For the prior example of $x = \frac{11}{20}$, in Fig 3.3, $b = 11$ and $a = 20$, and as can be seen in the figure, $u[n]$, $v[n]$, and $y[n]$ do have the expected period of 40. For the special case of $x = 0$, the output oscillates between V and $-V$ and the output spectrum consists of a pure tone at $\frac{f_s}{2}$. More complete results, which are independent of the integrator initial condition, $v[0]$, are provided in [10], [11], and [13]. In fact, it is shown in [11] and [13] that even if the DC input is an irrational multiple of V, the quantization noise will not be white and the spectrum at the output of the modulator will be discrete, consisting of tones.

Even for sinusoidal inputs, the quantization error is not white and strong tone components are observed in the output and the strength of the tone distortion components depends on the input amplitudes in a complicated way [15]. The tone structure present at the output of the modulator for very low DC or sinusoidal input amplitudes is often called idle channel noise.

One other point is worth noting. It can be shown that for comparator output levels of $\pm V$, the output of the integrator can have magnitude of at most $2V$ if the input to the modulator is bounded by $\pm V$ [15]. This is easily seen from (3.6) and (3.8) from which we have $v[n] = x[n-1] - (y[n-1] - v[n-1])$ which is $v[n] = x[n-1] - e[n-1]$. If we assume $|v[n-1]| \leq 2V$ then from the transfer characteristic of the 1 bit quantizer (Fig 2.2(c)) , $|e[n-1]| \leq V$ i.e. the quantizer is not overloaded. We then have, $|v[n]| = |x[n-1] - e[n-1]| \leq |x[n-1]| + |e[n-1]| \leq V + V = 2V$. Thus if $|x| \leq V$ and $|v[n-1]| \leq 2V$ then $|v[n]| \leq 2V$. This can be guaranteed by ensuring that $|v[0]| \leq 2V$, so that $|v[1]| \leq 2V$

and so on.

In practice, because of the significant tone structure present at the output of a first order ΔΣ A/D, it is rarely used in applications such as speech or audio where the presence of such tones is objectionable even if the oversampling ratio, $\frac{f_s}{2f_B}$, is high enough to provide a good overall SNR based on the linearized white noise model.

3.4 Higher Order ΔΣ Modulation

The fundamental ideas presented can be extended to create ΔΣ architectures in a variety of ways that provide different tradeoffs among resolution, bandwidth, circuit complexity, and modulator stability. Our discussion will include higher order, multi-bit, and multi-stage (cascaded) architectures, and band-pass architectures. In general, to obtain a performance improvement, most of these converters require analog circuits which need to be more complex and precise than those used in the 1st order ΔΣ modulator. Of course, the precision required must still be significantly less than the overall high conversion resolution. We begin our discussion with the standard second order modulator.

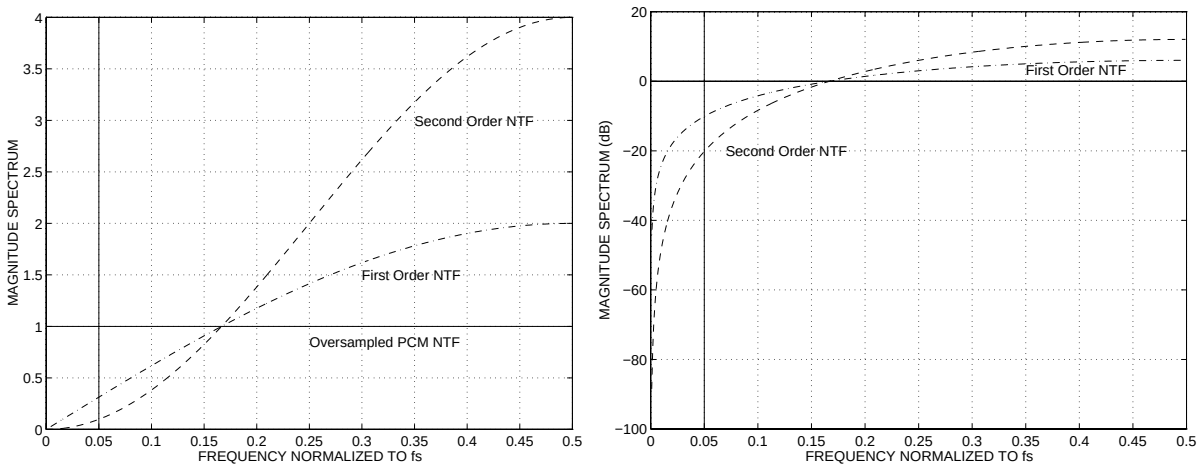
3.4.1 Second Order ΔΣ Modulation

Operation and Performance Modeling

The standard 2nd order delta-sigma modulator A/D is widely used. Unlike the 1st order modulator, the 2nd order modulator uses two discrete time integrators instead of just one and realizes $H_x(z) = z^{-1}$ and $H_e(z) = (1 - z^{-1})^2$ so that

$$Y(z) = X(z)z^{-1} + E(z)(1 - z^{-1})^2 \tag{3.12}$$

The 2nd order modulator noise transfer function (NTF) is shown in Fig 3.5 along with the NTF for the 1st order modulator. Note that compared with the 1st order ΔΣ NTF, the 2nd order NTF provides more quantization noise suppression over the low frequency signal band and more amplification of the noise outside the signal band. Compared with a 1st order ΔΣ, more noise power is *pushed* outside the signal band. A block diagram of the modulator is shown in Fig 3.6. Note that the structure now contains two integrators. The



(a) (b)

Figure 3.5: NTF for 1st and 2nd order delta-sigma modulators (a) Magnitude spectra on linear scale. For comparison, the oversampled PCM NTF, which is unity gain, is shown. (b) Magnitude spectra in dB. The vertical bar demarcates the extent of the signal band, f_B, where $f_B = 0.05 f_s$ in this example.

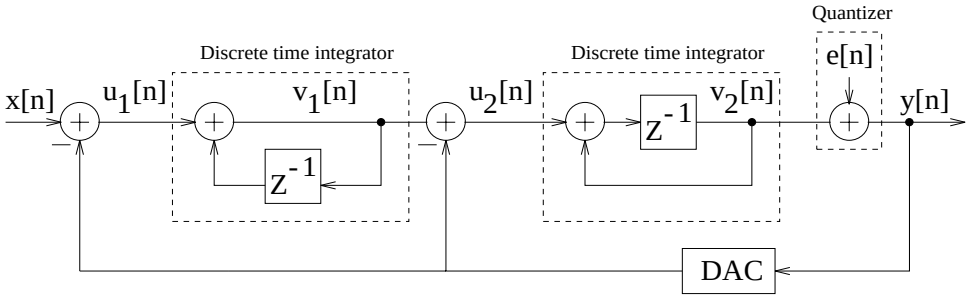


Figure 3.6: Second order delta-sigma modulator

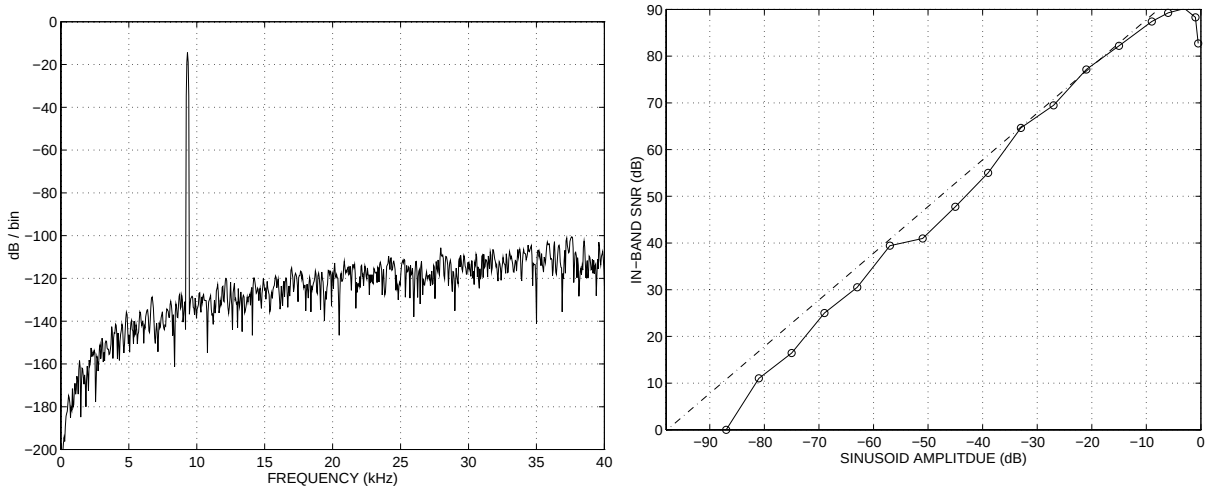
transfer function of the first one is $\frac{1}{1-z^{-1}}$ and that of the second one is $\frac{z^{-1}}{1-z^{-1}}$. Assuming the modulator output is filtered by an *ideal* low-pass filter, the linearized white noise model yields the following for the in-band SNR,

$$SNR = 10\,log(\sigma_x^2) - 10\,log(\sigma_e^2) - 10\,log\left(\frac{\pi^4}{5}\right) + 50\,log\left(\frac{f_s}{2f_B}\right)\,(dB) \tag{3.13}$$

Again, letting $\frac{f_s}{2f_B} = 2^r$, we obtain

$$SNR = 10\,log(\sigma_x^2) - 10\,log(\sigma_e^2) - 10\,log\left(\frac{\pi^4}{5}\right) + 15.05r\,(dB) \tag{3.14}$$

Thus, for every increment in r or for every doubling of the oversampling ratio, $\frac{f_s}{2f_B}$, the SNR improves by 15 dB or the resolution improves by 2.5 bits which is 1 bit better than



(a) (b)

Figure 3.7: Simulation of a 2nd order modulator (a) Power spectrum of the modulator output before decimation (b) in-band SNR vs amplitude of input sinusoid

the improvement achieved by a 1st order ΔΣ. Let us consider the example used previously where a 20 kHz audio band needs to be converted to 16 bits of resolution. The f_s needed by a 2nd order ΔΣ modulator using only a **1 bit** quantizer can be computed from (3.13) to be only 6.12 MHz in contrast with the 96 MHz needed by the 1st order ΔΣ ! This circuit speed is very reasonable in current CMOS technology.

Fig 3.7(a) shows the low frequency portion (0 to 40 kHz) of an FFT based power spectral density estimate of the output of a 2nd order modulator with a sinusoidal input frequency of 9.3 kHz and f_s of 6.20 MHz. The tall peak is, of course, the sinusoidal signal. Notice the noise shaping whereby the noise at lower frequencies is greatly attenuated. Such power spectra are often used to numerically calculate the in-band SNR. Here, the in-band SNRs are computed over the 20 kHz audio band. The SNR computation can be repeated for various sinusoid amplitudes to obtain a plot of in-band SNR vs amplitudes. This plot is shown in Fig 3.7(b). The dashed curve corresponds to the values which are obtained from the linearized white noise model SNR formula (3.13). The discrepancy between the two plots will be discussed later. Using the definition of dynamic range, D_R, it can be determined from the in-band SNR vs amplitude plot by looking at the amplitude value for which the SNR is 0 dB. The peak SNR is about 90 dB and the dynamic range, D_R, is about 88 dB which can be used in (2.5) to calculate an equivalent resolution, N, of 14.32 bits.

Qualitative Time Domain Behavior

In the time domain, referring to Fig 3.6, we have,

$$u_2[n] = v_1[n] - y[n]$$

$$v_2[n] = u_2[n-1] + v_2[n-1]$$

$$y[n] = \begin{cases} 1 & v_2[n] \geq 0 \\ -1 & v_2[n] < 0 \end{cases}$$

$$u_1[n] = x[n] - y[n]$$

$$v_1[n] = u_1[n] + v_1[n-1]$$

Figs 3.8(a) and (b) show plots of a DC input, $x[n] = 0.55$, and the resulting output, $y[n]$, of a 2nd order $\Delta\Sigma$ modulator. The input and output of the 1st integrator are $u_1[n]$ and $v_1[n]$ respectively, while the input and output of the 2nd integrator are $u_2[n]$ and $v_2[n]$ respectively. The "error" between the modulator input and output is $u_1[n]$ which again is not the quantization error given by $e[n] = y[n] - v_2[n]$. Looking at $u_2[n]$, we see that it is the difference between an integrated (or low-pass filtered) version, $v_1[n]$, of the modulator "error", $u_1[n]$, and the output, $y[n]$. Thus, $u_2[n]$ can be considered to be a more fine or accurate version of the modulator error. The signal which is quantized, $v_2[n]$, is an integrated version of the "fine error", $u_2[n]$. Consequently, $u_1[n]$, and $v_1[n]$ are analogous to the $u[n]$ and $v[n]$ of the 1st order modulator while $u_2[n]$ and $v_2[n]$ are more accurate representations of $u_1[n]$ and $v_1[n]$ and thus produce an output $y[n]$ which is more accurate than the output of a 1st order modulator. This should be clear from comparing Fig 3.8(c) with Fig 3.8(e) and Fig 3.8(d) with Fig 3.8(f). In comparing $y[n]$ of the 2nd order $\Delta\Sigma$ of Fig 3.8(b) to the $y[n]$ of the 1st order $\Delta\Sigma$ of Fig 3.3(b), the key point to note is that the distribution of 1's and -1's in Fig 3.8(b) is such that when averaged, the average provides a more accurate representation of the input than the corresponding average of the 1st order modulator output. In other words, for a given block of output samples, the 2nd order modulator uses its allocation of samples more efficiently to represent the input.

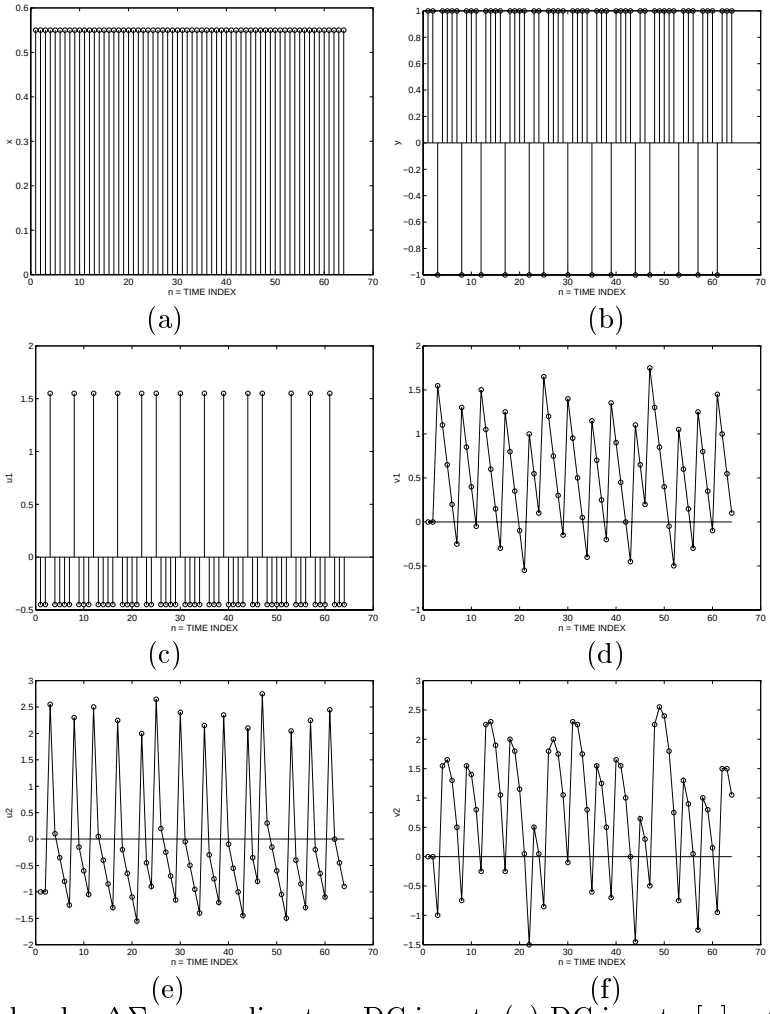


Figure 3.8: 2nd order ΔΣ responding to a DC input: (a) DC input $x[n] = 0.55$ (b) output $y[n]$ (c) "error" signal $u_1[n]$ (d) first integrator output $v_1[n]$ (e) more accurate "error" signal $u_2[n]$ (f) second integrator output $v_2[n]$

Implementation Imperfections

Compared with a 1st order ΔΣ, a 2nd order ΔΣ modulator contains an extra integrator. Assuming the same leakage factor, α, for both integrators, the degraded NTF is approximately $(1-\alpha z^{-1})^2$. The leakage factor of the second order integrators can satisfy somewhat less stringent requirements than those of the 1st order modulator integrators. One might expect this since, despite the NTF zeros being moved inside the unit circle, there are still two of them providing attenuation of quantization noise. The reduced requirement on the 2nd order modulator integrator leakage factors can also be seen if one calculates the in-band

noise power, σ_{ey}^2, from the $(1-\alpha z^{-1})^2$ NTF. The noise term which is inversely proportional to the oversampling ratio $\frac{f_s}{2f_B}$ is now divided by A^4 rather than A^2 as was the case for the 1st order modulator in (3.9). Consequently, a lower opamp gain can suffice for the 2nd order modulator.

The other parameter to consider is the gain of the 2nd integrator. From the linearized analysis, the STF and NTF poles are stable for integrator gains up to $\frac{4}{3}$ and so still allow a relatively wide variation from this point of view. Simulations indeed show that the integrator gains are relatively insensitive to deviations from their nominal values over a wide range of oversampling ratios [9].

Non-linear Behavior

Like the 1st order ΔΣ modulator, a 2nd order modulator may also display limit cycle oscillations [2], [10] and this is easily illustrated in a manner similar to that used for the 1st order modulator. The nature of these limit cycles has been investigated and unlike the 1st order modulator depend on the initial conditions of the integrator outputs [10], [12].

Most of the analyses e.g. [14], [15], [16] which provide an exact description of the spectrum of the quantization error and the modulator output for DC or sinusoidal inputs have been performed for 2nd order modulators using quantizers with 2 or more bits. In fact, an 1 bit quantizer used in a 2nd order modulator can become overloaded thereby making the analysis much more difficult. The quantizer is overloaded because the output of the second integrator can significantly exceed values of $\pm 2V$ [2] even with the modulator input bounded by the quantization levels $\pm V$. This is particularly true for large modulator inputs near the quantization levels. However, it has been determined from simulations that a modified 2nd order architecture [7] using an 1 bit quantizer can operate without the integrator outputs having to significantly exceed values of $\pm 2V$.

The spectral properties of overload noise have not been theoretically characterized in the literature but simulations reported in the literature [17], [18] have demonstrated that they significantly manifest themselves as harmonic distortion tones for sinusoidal inputs as well as significant tone components near $\frac{f_s}{2}$. As with the first order ΔΣ, idle channel tones may be observed for small DC or low amplitude tone inputs. According to the linearized model

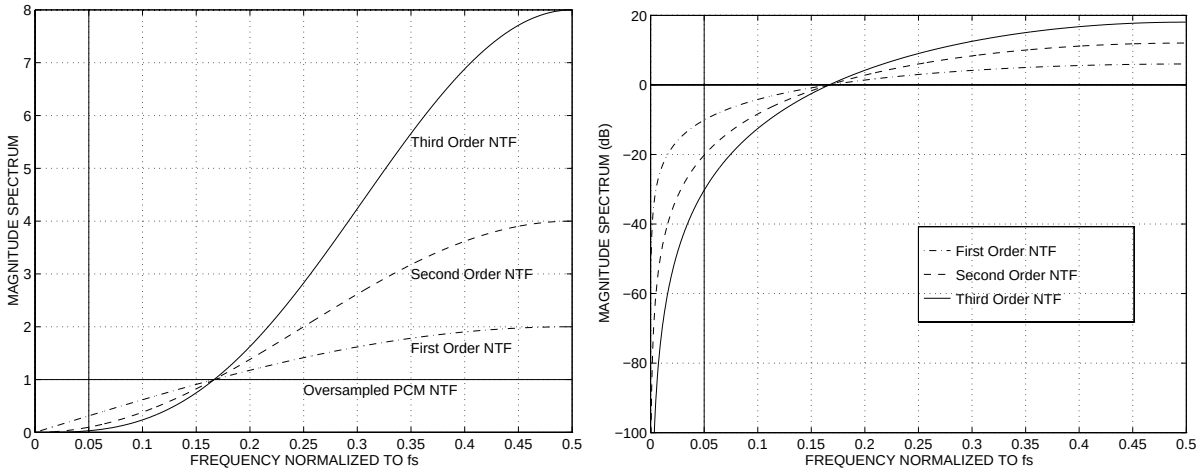
(3.13), σ_e^2 is fixed and the SNR should increase linearly with signal power σ_x^2. However, due to the presence of overload or idle channel tones, the SNR of a 2nd order modulator, using an 1 bit quantizer, increases linearly with signal power only over a certain range of signal power even though the modulator input may be between the quantization levels $\pm V$. More and more overload noise power is produced with increasing input values. Consequently, above a certain value of high signal power, the SNR will actually start to decrease when the increase in overload noise power dominates the increase in signal power. This can be seen in Fig 3.7(b) for input amplitudes greater than about -5 dB. On the other hand, as the input power becomes low, the SNR decrease is caused both by the decrease in signal power and by the presence of idle channel tone noise in the signal band. This can clearly be seen in Fig 3.7 for low amplitude values. According to (3.13) based on the linearized white noise model, the peak SNR and dynamic range should have been about 98 dB. Thus, idle channel tone problems, resulted in a 10 dB degradation of the dynamic range predicted by the linearized model while overload noise prevented the modulator from reaching the peak SNR predicted by the model.

Dithering techniques [18] will often break up tone structures including overload and idle channel tones thereby producing a smoother power spectrum output, a more linear dependence of SNR on signal power, and a dynamic range which is closer to the value predicted by (3.13).

Another factor which affects the in-band SNR is the sinusoid input frequency. This is particularly true if the sinusoid amplitude is such that strong harmonic distortion components are produced in the power spectrum output. In this case, the choice of a higher input frequency will result in these harmonics falling outside the 20 kHz bandwidth and not contributing to the in-band SNR. On the other hand, the choice of a lower input frequency will yield poorer values of in-band SNR because the harmonics will now fall inside the 20 kHz signal band.

3.4.2 Other Types of Higher Order $\Delta\Sigma$ Modulation

Delta-sigma converters realizing higher order NTFs achieve even higher resolution by pushing even more noise power outside the signal band. Alternatively, a lower sampling rate



(a) (b)

Figure 3.9: NTF for 1st and 2nd order delta-sigma modulators (a) Magnitude spectra on linear scale. For comparison, the oversampled PCM NTF, which is unity gain, is shown. (b) Magnitude spectra in dB. The vertical bar demarcates the extent of the signal band, f_B, where $f_B = 0.05 f_s$ in this example.

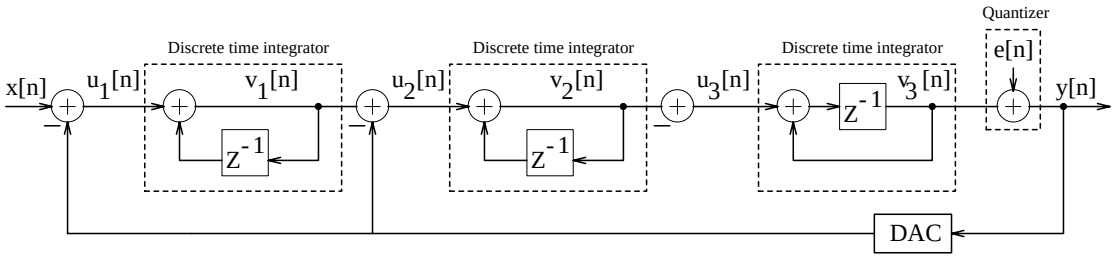


Figure 3.10: Third order delta-sigma modulator

can be used to achieve the same resolution for a given signal bandwidth. In this case, the speed requirements on the analog hardware is relaxed. An order L modulator based on a straightforward extension of the first order $\Delta\Sigma$ realizes a STF given by $H_x = z^{-1}$ and a NTF given by $H_e = (1 - z^{-1})^L$ which contains L zeros at $z = 1$ or at DC frequency on the unit circle. The magnitude spectra for a 3rd ($L = 3$) order NTF plotted on a linear scale and in dB are shown in Figs 3.9(a) and (b). Note that over the signal band, which is $0.05 f_s$ in the figure, the 3rd order NTF provides more attenuation of the quantization noise than the 2nd or 1st order NTFs and so is capable of pushing more noise power outside the signal band than the 2nd or 1st order modulators. The block diagram of a 3rd order ($L = 3$) modulator structure is shown in Fig 3.10. The ideal in-band SNR achieved by an Lth order

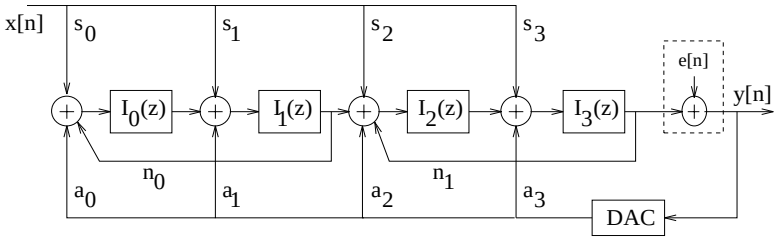


Figure 3.11: An example of a fourth order modulator topology

modulator is given by

$$SNR = 10\,log(\sigma_x^2) - 10\,log(\sigma_e^2) - 10\,log\left(\frac{\pi^{2L}}{2L+1}\right) + (20L+10)\,log\left(\frac{f_s}{2f_B}\right)\ (dB) \quad (3.15)$$

Thus, for every doubling of the oversampling ratio, this modulator provides an extra $(6L+3)$ dB of SNR or an extra $\left(L+\frac{1}{2}\right)$ bits of resolution. For the CD example being used throughout this chapter, a 3rd order modulator will need a f_s of 1.92 MHz to convert a 20 kHz band to 16 bits of resolution.

Modulator topologies are not constrained to always realize FIR NTFs as we have seen until now or to realize all the NTF zeros which provide attenuation at DC frequency. In fact, distributing the zeros over the signal band rather than placing them all at DC frequency can be more efficient in pushing quantization noise outside the signal band. Examples of higher order topologies are described in [19], [20], [21], [22]. One such fourth order topology described in [20] is shown in Fig 3.11 where $I_k(z)$ denotes the kth integrator. This structure realizes (2.7) in the form

$$Y(z) = X(z)\frac{B_x(z)}{A(z)} + E(z)\frac{B_e(z)}{A(z)} \quad (3.16)$$

Note that the STF and NTF are IIR transfer functions in this case. The feedforward coefficients s_k realize B_x which contains the Z domain zeros of the STF. The feedback coefficients n_k between every second integrator realize B_e which contains the zeros of the NTF. Finally the feedback coefficients a_k realize A which contains the poles of both the STF and the NTF. It can be shown that in order to implement (3.16) based on a $\Delta\Sigma$ modulator, which incorporates feedback, it is required for the z^0 coefficient in B_e and A to be equal to ensure a causal feedback loop (in other words to avoid non-computable delays, [5, pp. 308-309]). If the NTF is FIR i.e. if $A(z) = 1$ then the causality constraint will require the

z^0 coefficient of $B(z)$ to be 1. Note that the performance will be limited by the degree to which the analog coefficients a_k, s_k, and n_k, match their desired values. For the modulator to be useful, the degree of required matching should be significantly less than the overall resolution of the converter.

Higher order architectures also alleviate some of the tone problems mentioned earlier [2], [19]. The main difficulty with such higher order structures is that such modulators are only conditionally stable. Stability may for example depend on the input signal being kept below a certain value or on precise circuit matching needed to satisfy a stability criterion. Stability is often described in the sense of the quantizer not being overloaded. This is useful because any higher order modulator structure can be transformed into an equivalent modulator in the form of the 1st order modulator of Fig 3.1 with the integrator being replaced by a general "loop filter" $H(z)$. If the quantizer is not overloaded, its input is bounded by $\pm 2V$ and this implies that the loop filter is also operating in a stable manner and all internal signals will be bounded. Stability is, in general, difficult to determine for a modulator using an 1 bit quantizer. One reason is the difficulty in characterizing the gain of the 1 bit quantizer. The gain of the 1 bit quantizer of Fig 2.2(c) is variable – it depends on the quantizer input. The smaller the quantizer input the larger is its gain (e.g. if the input is zero, the output is V and the gain is infinite) while the larger the quantizer input the smaller is its gain (e.g. if the input is infinite, the output is V and the gain is zero). Another way of looking at this is to observe that if one attempts to linearize the quantizer transfer curve of Fig 2.2(c) by trying to fit a straight line to the curve, the correct slope of the straight line is arbitrary. As we have seen earlier, the input to the quantizer clearly changes with time even if the input to the modulator is a DC signal. The manner in which the quantizer gain changes over time will also depend on the type of input applied to the modulator. Consequently performing a linear system analysis of the modulator signal and noise transfer functions in terms of modulator parameters is inadequate because the poles of the transfer function which determine stability depend on a time varying and input dependent quantizer gain. An attempt to characterize the quantizer gain more accurately for DC and sinusoidal inputs has been made for several modulators [23].

The phenomenon of limit cycle oscillations is also connected to stability. This is because

the structure of limit cycles may be such that the amplitude of internal modulator variables are large causing the quantizer to overload. The frequency of such a deleterious limit cycle oscillation can correspond to the point on the unit circle where the modulator transfer function pole crosses the unit circle into the unstable region. Properties of limit cycles in the context of stability have been investigated in [12]. It has been pointed out [24] that a limit cycle which corresponds to the modulator transfer function pole moving from the inside to the outside of the unit circle may not necessarily result in unstable behavior in the long term provided integrator outputs do not saturate before stability is restored. Suppose an unstable limit cycle, corresponding to poles moving outside the unit circle, results when the quantizer gain is too high, i.e. when the input to the quantizer is small. If this is the case, growing signal values in the modulator which result from this instability will eventually increase the input to the quantizer thereby reducing its gain and so moving the poles back inside the unit circle. One way to guarantee stability would be to reset the integrators if it was detected, by additional circuitry, that their values were becoming too large. However, this approach may cause a significant decrease in the SNR [25]. Similarly, allowing integrator outputs to clip or saturate may also cause degradations in the SNR performance. In particular, low frequency limit cycles which may introduce distortion components in the signal band may occur (albeit with reduced values of integrator output) and may persist for a long time thereby reducing the SNR [24]. An alternative approach uses local feedback loops in an attempt to gracefully return integrator outputs to their normal operating region [25]. Since the linearized transfer function of the system is modified due to the local feedback loops, the effect of these loops is cancelled in the digital domain.

Several quantitative criteria have also been proposed to characterize stability. The l_1 norm criterion [15], [26], [27] relates the sum of the magnitudes of the modulator NTF impulse response coefficients, the number of bits in the quantizer, and the modulator input level to the no overload stability requirement. This is a sufficient condition but not necessary condition for stability and has unfortunately been found to be too conservative for practical use. An ad-hoc stability criterion which has been proposed [19] and found to be useful [22] is to design the NTF to possess less than 2 to 6 dB of out-of-band gain.

3.5 Multi-bit ΔΣ Modulation

Until now we have assumed that the quantizer and DAC inside our ΔΣ modulator were 1 bit devices. However, converters using a multi-bit internal quantizer offer more potential resolution from the internal quantizer. A 2nd order multi-bit ΔΣ converter would look exactly the same as the modulator shown in Fig 3.6 except that $e[n]$ in the figure would be the model for a N bit quantizer instead of an 1 bit quantizer, and the DAC would be a N bit DAC instead of an 1 bit DAC. The use of a multi-bit quantizer affects the σ_e^2 term in the expressions for the SNR, where each additional bit used in the quantizer will yield a 6 dB improvement in the SNR. Using (2.3) without the approximation, it is easy to see that if a 5 bit internal quantizer is used instead of an 1 bit quantizer, a 30 dB improvement in SNR is possible. Alternatively, the sampling frequency can be reduced by a factor of 4 while keeping the resolution the same. For our CD example where 16 bit resolution is desired for a 20 kHz signal bandwidth, a 2nd order modulator using a 5 bit internal quantizer can use a f_s of 1.53 MHz rather than the f_s of 6.12 MHz needed by a modulator using an 1 bit quantizer.

The behavior of multi-bit ΔΣ systems more closely follow that predicted by the linearized model (in the extreme case, if the quantizer has an infinite number of bits, there is no non-linearity). Consequently, the stability of higher order modulators using multi-bit quantizers is generally more accurately predicted. Another way of viewing enhanced stability is to consider the gain of the multi-bit quantizer. If a midtread multi-bit quantizer is used, its gain is relatively close to one for most output values (even for zero input, a midtread quantizer will have a zero output and hence unity gain unlike the midriser 1 bit quantizer which has infinite gain for a zero input). This is because a straight line drawn through the multi-bit quantizer transfer characteristic (e.g. Fig 2.2(b)) can no longer be arbitrarily drawn. Of course, if for any reason the quantizer does start to overload, its gain will start to deviate more and more from unity. Even though the l_1 stability criterion mentioned earlier may be too conservative, it may allow one to obtain an initial idea about the number of bits needed for stable operation for a given NTF and input signal level.

Modulators using multi-bit quantizers also display less of the tone problems associated with the 1st and 2nd order delta-sigma converters using an 1 bit internal quantizer. The

multi-bit output also complicates the digital low-pass filter hardware following the modulator because for multi-bit processing, the filter requires multi-bit hardware multipliers. However, the main disadvantage is that the multi-bit DAC cannot be easily fabricated in VLSI with sufficient linearity needed for high resolution conversion. Various techniques in the literature have been proposed to reduce the linearity required for the DAC. For example, in [28] and [29] the multi-bit DAC is calibrated and the information obtained from this process is used to digitally correct for the multi-bit DAC non-linearity. This is done by taking each possible input codeword of the DAC, converting it to an analog quantity, and then actually using a single bit delta-sigma A/D modulator system to digitize the multi-bit DAC's analog output. Thus a digital representation of the DAC's input/output relationship has been obtained and can be stored in a RAM lookup table for subsequent use in the digital correction process. The approach in [30] was to use a N quantizer but only a perfectly linear one bit DAC in the feedback path. The input to the one bit DAC is the most significant bit (MSB) of the N bit quantizer. The LSBs of the quantizer are filtered with a digital filter and combined with N bit output of quantizer to remove any effect of using only a 1 bit DAC in the feedback path.

3.6 Multi-stage (cascaded) ΔΣ Modulation

Higher order NTFs can also be created by cascading independent modulator stages. This cascading does not adversely affect the stability of the overall modulator provided the individual stages are stable. They may also suffer less of the tone problems than a 1st or 2nd order delta-sigma alone [2], [15]. An example of a 2nd order modulator obtained by cascading two 1st order modulators is shown in Fig 3.12. The signal $x[n]$ is the input to the first modulator in the cascade but the quantization error, $e_1[n]$, of the first modulator is used as the input to the second modulator. Finally, the outputs of the first stage and second stage modulators are added in the digital domain after passing through a digital delay (z^{-1}) and a digital differentiator ($1 - z^{-1}$), respectively. From the Z domain analysis of the linear system model with the DACs replaced by unity gains, we have,

$$Y_1(z) = X(z)z^{-1} + E_1(z)(1 - z^{-1}) \tag{3.17}$$

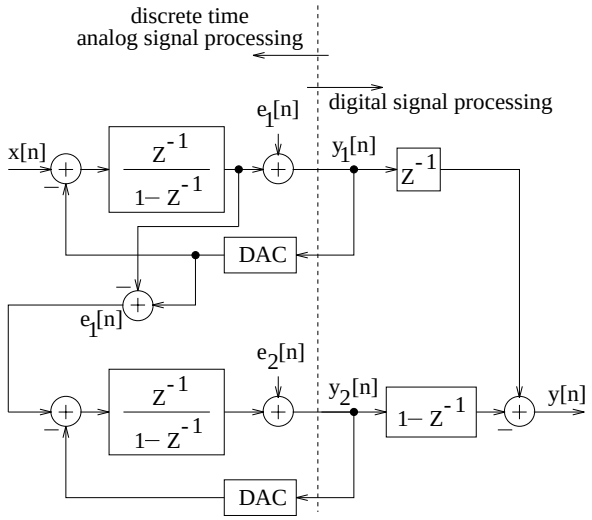


Figure 3.12: "1-1" cascade: a 2nd order modulator from a cascade of two 1st order modulators

$$Y_2(z) = E_1(z)z^{-1} + E_2(z)(1 - z^{-1}) \tag{3.18}$$

The output is computed as $Y_1 z^{-1} - Y_2(1 - z^{-1})$. This sum results in a cancellation of the first order noise term $E_1(z)$ to produce the overall output,

$$Y(z) = X(z)z^{-2} - E_2(z)(1 - z^{-1})^2 \tag{3.19}$$

Except for the sign on the noise that is irrelevant and an extra delay experienced by the input, the modulator realizes the same output as the standard 2nd order ΔΣ modulator. One advantage of using this structure over the 2nd order modulator is the fact that the quantizer in either of the 1st order modulator sections will never overload for $x[n]$ bound by $\pm V$. However, the cascaded structure requires matching between the analog and digital transfer functions as well matching among the D/A output levels among various stages [2], [35]. In fact, mismatch effects and integrator leakage can lead to the propagation of unshaped or poorly shaped noise from an earlier section to the final output [36], [37]. Assume there are circuit imperfections in the "1-1" cascade of Fig 3.12, such that the transfer function of the integrator in the first section is $\frac{gz^{-1}}{1-\alpha z^{-1}}$ instead of $\frac{z^{-1}}{1-z^{-1}}$. Even if we assume the integrator in the second stage is ideal, the output is then

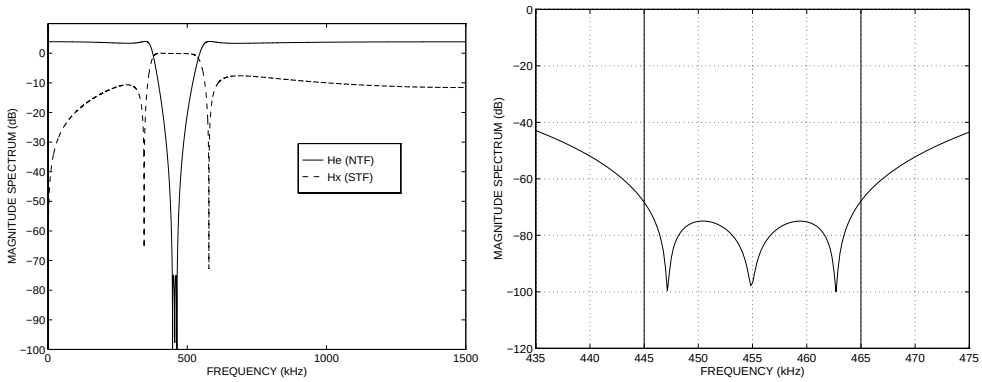
$$Y(z) = \frac{X(z)gz^{-2}}{1-(g-\alpha)z^{-1}} + \frac{(1-\alpha)z^{-2}E_1(z)}{1-(g-\alpha)z^{-1}} + \frac{(g-\alpha)z^{-2}E_1(z)(1-z^{-1})}{1-(g-\alpha)z^{-1}} - E_2(z)(1-z^{-1})^2 \tag{3.20}$$

The first term of the equation contains the signal which is no longer a pure delay but will have a ripple determined by the factor $\frac{g}{1-(g-\alpha)z^{-1}}$. However, since the signal is oversampled, this additional signal transfer function will mostly be flat at lower baseband frequencies. Ignoring the ripple due to the factor $\frac{1}{1-(g-\alpha)z^{-1}}$, the second term in (3.20) is the unshaped noise from the first stage while the third term in (3.20) is the first order shaped noise from the first stage, while the fourth term in (3.20) is the desired second order shaped noise term. Clearly, for large imperfections, the unshaped noise term might dominate the noise term subject to 2nd order noise shaping. If more than two stages are cascaded, the cumulative effects of such quantization error leakage effects will yield diminishing returns in the performance improvement. Architectures using only first order modulators have been realized [38] as have architectures using second order modulators [39], [40], [41]. A comparison of some architectures can also be found in [37]. A multi-stage architecture has also been considered [42] for realizing higher order ΔΣ modulators which place some NTF zeros away from DC. Finally, note that due to the addition of various single bit intermediate outputs, the architecture has a multi-bit final output which complicates the decimation filter hardware.

3.7 Band-pass ΔΣ Modulation

So far we have assumed that the sampling frequency f_s is much greater than the Nyquist rate which is twice the highest frequency component in the input signal. For low-pass signals, the highest frequency component is also the signal bandwidth f_B. If a signal with bandwidth f_B is band-pass but is located at a center frequency, f_c, its highest frequency is $f_c + \frac{f_B}{2}$. If f_c is large, choosing f_s to be much greater than the highest frequency will lead to an unreasonably large f_s and does not take advantage of the band-pass nature of the signal. Band-pass delta-sigma modulation [43] allows high resolution conversion of band-pass signals if f_s is much greater than the signal bandwidth f_B rather than the highest signal frequency. Band-pass ΔΣ modulators can be used in AM digital radios [43] or receivers for digital cellular mobile radios [47].

Unlike low-pass ΔΣ modulators which realize NTF zeros at DC or low frequencies on



(a) (b)

Figure 3.13: Band-pass delta-sigma transfer functions: (a) Shown are magnitude spectra in dB of H_x (the STF, dashed plot) and H_e (the NTF) (b) Shown is a closeup of the magnitude spectrum in dB of H_e

the unit circle of the Z plane, band-pass modulators have NTFs which realize zeros or notches in the signal band of interest, $[f_c - \frac{f_B}{2}, f_c + \frac{f_B}{2}]$. Consequently, quantization noise, which occurs over the signal band, is attenuated and noise power is pushed outside this band. Regardless of where the signal band is centered, the smaller the signal band, f_B, relative to the sampling frequency, f_s, there is less in-band noise power for a given NTF. Noise outside the signal band can be attenuated with a digital decimation filter and so high resolution conversion is possible for large oversampling ratios $\frac{f_s}{2f_B}$. The modulator STF and the decimation filter will typically have a band-pass characteristic, providing unity gain over the signal band.

As an example, the NTF and STF magnitude spectra for the design in [44] are shown in Fig 3.13(a). In this example, the signal band is centered about a center frequency of $f_c = 455$ kHz, $f_s = 3$ MHz , $f_B = 20$ kHz so the oversampling ratio is 75. Fig 3.13(a) shows the magnitude spectra in dB of H_x (the STF) and H_e (the NTF). Fig 3.13(b) shows a closeup view of the magnitude spectrum in dB of H_e over the signal region The vertical bars delineate the 20 kHz signal band centered around 455 kHz. The NTF is 6th order, and as should be clear from Fig 3.13(b), contains 3 notches or zeros, which provide attenuation of the quantization noise, over the 20 kHz signal bandwidth (the other 3 zeros are complex conjugates of these). The STF is band-pass and has minimal ripple and is approximately linear phase in the signal band. Band-pass converters employing Lth order modulators display a SNR performance which improves at the rate of $(3L+3)$ dB per octave increment

with the oversampling ratio $\frac{f_s}{2f_B}$ [44].

Band-pass modulators can be implemented using modulator topologies such as the one in Fig 3.11 or those in [19] or [44]. A multi-stage (cascaded) band-pass architecture has been proposed in [48] for the case where the center frequency occurs at $\omega_c = \frac{\pi}{2}$.

3.8 Decimation for ΔΣ Modulation

This thesis will not discuss in great detail the various decimation structures used for the digital decimator following ΔΣ modulators. However, at this point, let us digress to make a few general comments. The closer the decimation filter approximates an ideal filter i.e. the higher the filter order, the better is the out-of-band quantization noise suppression and higher is the SNR. Of course, for a given modulator, there will be a point beyond which increasing the filter order will lead to insignificant improvements in the SNR. Also, the more efficient a ΔΣ modulator is in pushing in-band quantization noise power outside the signal band, the decimation filter needs to be more efficient in attenuating the out-of-band quantization noise to the extent that it becomes insignificant compared to the in-band noise. Consequently, for a given SNR performance after filtering, a more efficient ΔΣ modulator will require a higher order filter. For ΔΣ modulators with large oversampling ratios, the downsampling factor D needed to bring the sample rate down to the Nyquist rate after the low-pass filter can be rather high. A multi-stage decimation scheme is more efficient in terms of hardware (or computational) complexity when D is large [49]. The signal is first filtered with a low-pass filter with a somewhat relaxed transition band and stop-band attenuation specifications. It is then downsampled to an intermediate sampling rate. It is next filtered again with a low-pass filter whose specifications are tighter than the first stage filter specifications. Downsampling again follows the second stage filtering and so on until the signal is downsampled to the final desired rate. Since none of the filters in a multi-stage decimator will be ideal, there will always be some aliasing of out-of-band quantization noise into the signal band after each stage of downsampling. The extent to which this occurs will depend on the specifications for the decimation filters, the number of stages of decimation, the decimation factor of each stage, and the ΔΣ modulator NTF.

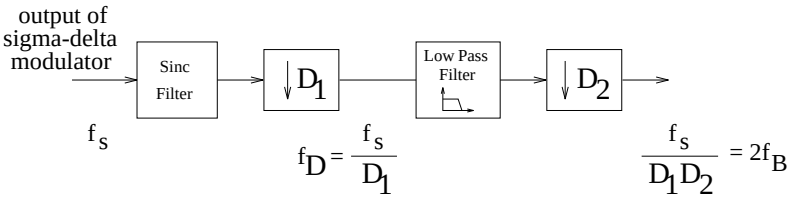
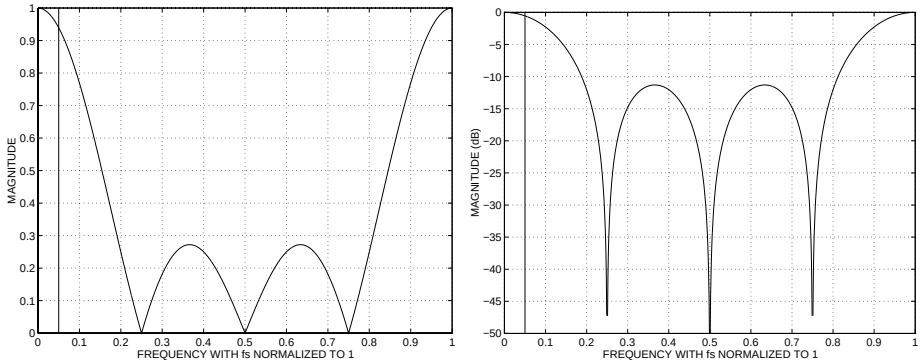


Figure 3.14: Two stage decimator with *sinc* filter as the first stage. $D = D_1 D_2$



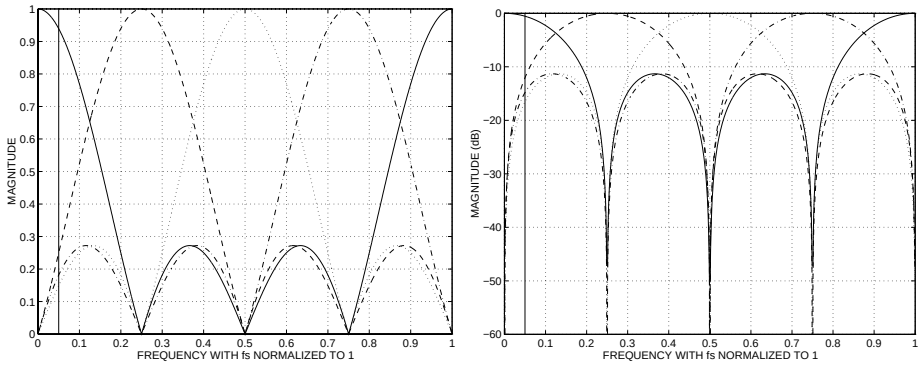
(a) (b)

Figure 3.15: Magnitude response of *sinc* filter $H(z) = \frac{1}{K}\frac{1-z^{-K}}{1-z^{-1}}$ with $K = 4$ (a) linear scale (b) dB scale. Frequency scale normalized to $f_s = 1$.

It turns out that for modulators realizing NTFs of the form $(1 - z^{-1})^L$, simple *sinc* filters can be used for the first stage of decimation without incurring a significant price in SNR loss due to aliasing of out-of-band quantization noise [2], [50]. A two stage decimator where the first stage is a *sinc* filter is shown in Fig 3.14.

The first stage downsampling factor is D_1 and the second stage downsampling factor is D_2. If the ΔΣ output is decimated all the way to the Nyquist rate, we have $\frac{f_s}{D_1 D_2} = 2f_B$. The impulse response of a Kth order *sinc* filter is $h[n] = 1$, for $0 \leq n \leq K - 1$ and $h[n]$ is zero otherwise. The transfer function is $H(z) = \frac{1}{K}\frac{1-z^{-K}}{1-z^{-1}}$ and is henceforth referred to as $sinc(K)$. The frequency response has nulls at multiples of $\frac{f_s}{K}$. A plot of the magnitude response of $H(z)$ for $K = 4$ is shown in Fig 3.15. The vertical bar is located at a signal bandwidth which for the sake of example is $0.05 f_s$.

Downsampling the ΔΣ modulator output after the $sinc(K)$ filter by a factor D_1 will result in out-of-band quantization noise being aliased into the signal band after attenuation by the $sinc(K)$ filter because the $sinc(K)$ filter does not have infinite out-of-band attenu-



(a) (b)

Figure 3.16: Magnitude response of $sinc(K)$ filter $H(z) = \frac{1}{K}\frac{1-z^{-K}}{1-z^{-1}}$ with $K = 4$ along with alias components (a) linear scale (b) dB scale

ation. Let the sampling frequency after the first stage of decimation, $\frac{f_s}{D_1}$, be f_D. If $Y(f)$ is the spectrum (here referred to by the continuous time frequency variable rather than by the Z transform variable) of the $\Delta\Sigma$ modulator output, its spectrum, $Y_{d1}(f)$ after filtering with a $sinc$ filter transfer function $H(f)$ and downsampling by a factor of D_1 is [5, pp. 101-105]

$$Y_{d1}(f) = \frac{1}{D_1}\sum_{k=0}^{D_1-1} Y(f - kf_D)H(f - kf_D) \tag{3.21}$$

where the $k \neq 0$ terms are the alias components. The shaped quantization noise at the output of the $\Delta\Sigma$ modulator will be weighted by the original filter response and the alias component responses before being added up to produce the overall noise component after downsampling. The noise will be weighted by the $sinc$ filter response from the region $[f_D - f_B, f_D]$, corresponding to $k = 1$, and from regions at multiples of f_D i.e. from $[kf_D - f_B, kf_D]$ for $k > 1$ before being aliased into the signal band $[0, f_B]$. Consequently, the nulls of the $sinc$ filter are chosen to occur at f_D and its multiples so that the out-of-band noise suffers a great deal of attenuation from the nulls of the filter before being aliased into the signal band. With the above choice of $sinc$ filter null locations, we have $K = D_1$. The noise contributions from alias components suffering attenuation from $H(f)$ nulls at higher frequencies will be less than those suffering attenuation from nulls at lower frequencies. The magnitude response of the alias components of the $sinc(K)$ filter are shown in Fig 3.16 for $D_1 = K = 4$. The solid line is the original $sinc(K)$ response and the other curves are the responses of the alias components. Note the contributions in the

signal band (to the left of the vertical bar) from the alias components. The attenuation of the $sinc(K)$ sidelobes does not change significantly with K but the attenuation of the sidelobes can be increased by using a higher order $sinc(K)$ transfer function of the form, $\left(\frac{1}{K}\frac{1-z^{-K}}{1-z^{-1}}\right)^{L+1}$ which is henceforth called $sinc(K)^{L+1}$. Several criteria guide the choice of $K = D_1$, and L. For a given L, if D_1 is is too large then there will be a large number of frequency components which alias into the signal band but the components corresponding to larger k in (3.21) will be attenuated more than those corresponding with smaller k. The relationship between $K = D_1$ and SNR degradation due to noise which aliases into the signal band is provided in [2]. There it is shown that for the Lth order modulator with NTFs of the form $(1 - z^{-1})^L$, the $sinc(K)^{L+1}$ filter are optimal in the sense that for small enough D_1, the SNR degradation is negligible. In fact, it is shown that the SNR does not degrade by more than 0.15 dB when $f_D = \frac{f_s}{D_1}$ is at least 4 times the Nyquist rate i.e. when D_1 is less than the oversampling ratio by at least a factor of 4. The value of $D_1 = K = 4$ has been chosen for the above examples only for easy visualization of the various plots. In practice, the oversampling ratio will be relatively large and so even if $D_1 = K$ is less than 4 times the oversampling ratio, $D_1 = K$ can still be relatively large. Another issue with the $sinc(K)^{L+1}$ filter is that the frequency response will contain a slight droop over the signal band. This is clear in Fig 3.15 where at the edge of the signal band the filter gain is less than 1 (less than 0 dB).

It is shown in [50] that as long as D_1 is at least four times smaller than the oversampling ratio, this droop is less than 1 dB for $L = 0$ i.e. for the $sinc(K)$ transfer function. Of course, for larger L, this droop will increase. This droop will need to be compensated in subsequent filter stages and so should be less than about 3 dB [2] for easy compensation. One of the main motivations of using $sinc(K)^{L+1}$ filters is that their implementation is simple because all the filter coefficients are 1 and thus no multipliers are needed. Efficient recursive implementations are also possible [2], [51]. It is desirable that the filter response of the low-pass filters be linear phase so as not to produce phase distortion in the signal. The $sinc(K)^{L+1}$ filter does possess linear phase and the low-pass filters of subsequent stages are usually designed to be FIR filters to provide linear phase.

For $\Delta\Sigma$ modulators which have NTFs that are not of the form $(1-z^{-1})^L$, the $sinc(K)^{L+1}$

may not provide the out-of-band quantization noise attenuation needed for the SNR to not be significantly different from the ideal in-band SNR. Although multi-stage decimation [49] using standard filter design techniques [52] can be used for these cases, design methods may also attempt to reduce the decimation filter complexity by optimizing the filter order with respect to the modulator noise transfer function [53], [54]. The use of IIR filters for decimation has also been considered in [55].

Chapter 4

Amplitude Slicing

This brief chapter introduces the notion of slicing in the context of A/D conversion. This is done by considering PCM A/D conversion and using two examples of PCM conversion to illustrate the slicing of amplitudes to perform A/D conversion. The idea is to break the conversion process so that one or more lower resolution ADCs can be used to convert different portions of the amplitude value. The block diagram of a pipelined A/D converter based on [56] is illustrated in Fig 4.1. Here the conversion of an amplitude sample v_{in} to N bits is done in K successive stages where an analog quantity is quantized in the ith stage using a N_i bit ADC. First v_{in} is quantized with a N_1 bit ADC. Let this value be Q_1 which can be converted to an analog quantity and used to obtain the "residue" (or remainder) $v_{in} - Q_1$ used as input by the next stage. The residue is amplified before quantization by

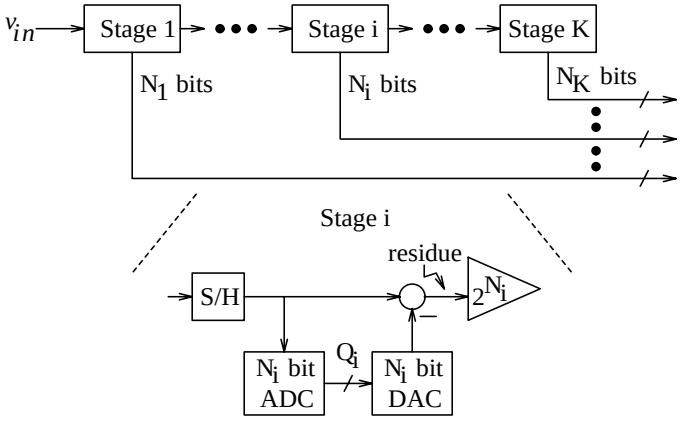


Figure 4.1: Block diagram of a pipelined ADC

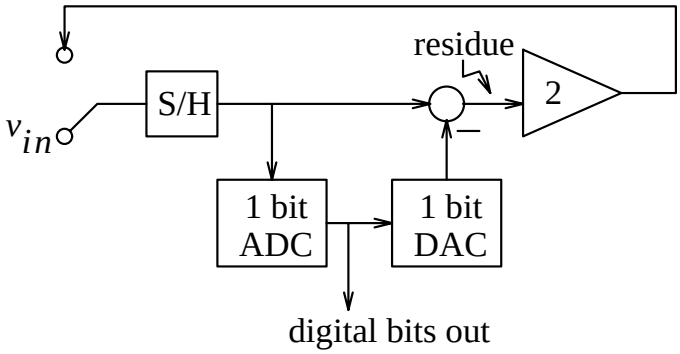


Figure 4.2: Block diagram of an algorithmic ADC

the next stage to make use of the full dynamic range of the quantizer of the next stage. In practice the sample and hold (S/H), residue computation and gain are performed by one S/H amplifier. Although the conversion appears to occur in a serial fashion and appears to require K clock cycles, pipelining of the computation allows a higher throughput of 1 clock cycle after a fixed delay. We observe that each sample amplitude is essentially sliced into various residues which are quantized by lower resolution converters. Quantizing a sample using an algorithmic ADC is similar and is shown in Fig 4.2. In some sense, this is a pipelined conversion performed in a pure serial manner, with a single 1 bit ADC, rather than in a pipelined fashion with multiple ADCs. However, the computation is again sliced into the quantization of different amplitude residues, each performed with a lower resolution ADC.

Although we have considered only two examples of PCM conversion, one can think of other types of converters as also performing a sort of amplitude slicing. For example, a flash converter uses 2^N comparators or 1 bit ADCs to compare v_{in} with 2^N reference levels. The 1 bit ADC outputs are combined to generate a thermometer coded output which is subsequently decoded to N bits. Note that each 1 bit ADC provides information about a different portion of the amplitude range of v_{in}.

Amplitude slicing is not directly applicable to $\Delta\Sigma$ converters in that it does not make sense to speak of $\Delta\Sigma$ A/D conversion for individual amplitude samples. Of course, the PCM quantizer in the $\Delta\Sigma$ modulator performs individual amplitude conversions based on amplitude slicing. In some sense, one might consider cascaded $\Delta\Sigma$ modulation to incorporate amplitude slicing to realize a higher order NTF. This is because the residue of an

internal quantized signal, of one stage is used as an input to an independent following stage and the outputs of the stages are combined to provide an overall output required for proper noise cancellation.

Chapter 5

Parallel Frequency and Time Sliced PCM Conversion

5.1 Introduction

We now describe and analyze the performance of frequency and time sliced PCM systems. Two frequency slicing architectures and a time slicing architecture are considered. The first is a standard sub-band coder structure where A/D converters are inserted as the coders. This A/D structure has been described in [57]. The second is a direct multi-band structure where no downsampling and upsampling are performed. This structure is motivated by this thesis's work on frequency sliced delta-sigma systems. In addition, a time sliced PCM architecture is examined. This architecture has been known for quite some time [58]. Although the sub-band coder frequency sliced and the time sliced architectures have been proposed in the literature, their quantization noise performance have not been directly compared or examined within a common framework. The thesis shows that that as the number of parallel channels is increased, the performance of the frequency and time sliced architectures do not improve in the same way. We analyze the quantization noise performance of these sytems using some results in the literature and explain the source of this performance difference. Simulation results are used to confirm the analysis.

It is important to note that the analysis and results of this chapter are performed within a **no overlaod framework**. This means that gain factors in the system are chosen in such a

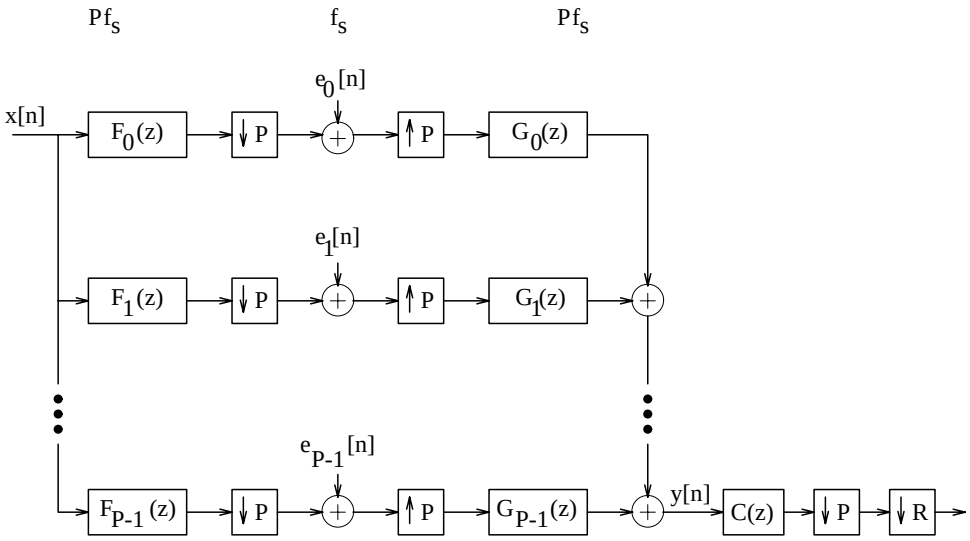


Figure 5.1: Block diagram of a sub-band coder PCM ADC

way such that the quantizers are never overloaded. This assumption is made for two reasons: (1) from the point of view of analysis, it is easy to compare the performance of various systems with internal quantizers producing the same quantization noise powers. Also, if one allows the converters to overload (even though this may not occur often statistically), the white noise assumptions about the quantization error process is less valid. (2) from a practical point of view, if the circuits are allowed to clip, they may produce other non-linearities in addition to the quantization.

5.2 Frequency Sliced PCM Conversion

5.2.1 Frequency Sliced Sub-band Coder Structure

Frequency sliced PCM conversion involves slicing the signal's frequency spectrum into various portions and quantizing each portion with independent ADCs whose outputs are recombined. We first examine a sub-band coder structure where ADCs are inserted as the coder [57]. Fig 5.1 shows a block diagram of a P channel sub-band coder PCM converter. Let the signal bandwidth be f_W. The oversampling ratio of the signal with respect to the Nyquist rate for the sampling frequencies at which the quantizers operate is $R = \frac{f_s}{2f_W}$. The signal is sampled at a rate Pf_s and then filtered with a bank of discrete-time (sampled data)

analog analysis filters $F_p(z)$. The output of each filter is downsampled by a factor of P to a sampling rate f_s. The ADC for each channel quantizes the resulting signal at this rate. The rates at which various blocks operate are shown on the top of the figure. The quantizer in the pth channel is modeled by an additive error or noise source $e_p[n]$. As is often done for the analysis of A/D conversion systems, we assume that the quantization error process in each channel is wide sense stationary (W.S.S.) and white. Each quantizer will be assumed to inject noise with a variance of σ_e^2 into its respective channel. After quantization, the quantized signals are upsampled by a factor of P and are filtered with a bank of digital synthesis filters $G_p(z)$ before all the channel outputs are recombined. After recombination, the output must be decimated by a factor of $R \times P$ to bring the sampling rate to the Nyquist rate $2f_W$. The block $C(z)$ represents some appropriate decimation filter. In general we will assume f_s is normalized to 2π and so we will use f_W and $\omega_W = \frac{2\pi f_W}{f_s}$ interchangeably.

Although the fundamental quantization noise analysis is provided in [59], we review the material here to point out implications for the system SNR performance as the number of channels is increased and to make comparisons with the time interleaved method to be discussed later. Defining, $W = \exp(-j\frac{2\pi}{P})$, the output of the system prior to decimation is

$$Y(z) = \sum_{l=0}^{P-1} X(zW^l) \left(\frac{1}{P}\right) \sum_{p=0}^{P-1} F_p(zW^l) G_p(z) + E(z) \tag{5.1}$$

where the overall quantization error component, $E(z)$ is given by,

$$E(z) = \sum_{p=0}^{P-1} E_p(z^P) G_p(z) \tag{5.2}$$

and the l - sum is the summation of the alias components while the p - sum is the summation of the P channel responses. We will assume that the F_p and G_p filters are designed to provide perfect reconstruction of the signal component. However, note that to ensure that signal component in $Y(z)$ is reconstructed with a gain of 1, there must be a gain of P through the path of each channel i.e. the cascade of F_p and G_p should provide a gain of P. The distribution of this gain between F_p and G_p (for example with F_p and G_p having a gain of $\sqrt{P}$ each) depends on the maximum amplitude produced at the output of F_p and has implications for the quantization noise performance of the A/D converter especially in the context of our **no overload framework**.

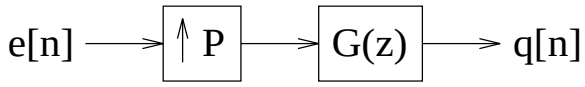


Figure 5.2: Upsampling followed by filtering

Consider the maximum amplitude at the output of F_p. If a white uniformly distributed random process, is used as the input to a brickwall filter with passband $\frac{1}{P}$ of the total bandwidth, the power in the output process will be reduced by a factor of P. However, the maximum amplitude of the output will *not*, in general, be reduced by a factor of $\sqrt{P}$. The reduction in the maximum amplitude can be reduced by a factor of $\sqrt{P}$ only if the output of the filter is uniformly distributed. However this will not be the case in general. If we assume that the random process input to the linear filter is indentically and independent distributed (i.i.d.) and distributed uniformly between $-\frac{\Delta}{2}$ and $\frac{\Delta}{2}$, the output probability density function (p.d.f.) will be convolutions of the input p.d.f.s scaled by the impulse response coefficients (in an extreme case, if all the coefficients were unity and the filter had infinite length, the output p.d.f. would be Gaussian by the central limit theorem). Theoretically, the maximum amplitude range after filtering will always increase. However, the probability of larger amplitude values occurring will decay. Consequently, depending on the filter, the amplitude range observed in simulations after filtering, could be somewhat less than before filtering.

If we assume that this reduction in amplitude of the more probable amplitudes at the output of F_p is not significant, the filter F_p should have a gain less than 1 so the quantizer following it is not overloaded for our **no overload framework**. Consequently, the digital filters G_p must have a pass band gain of P to provide the gain of P required to reconstruct the signal properly. This has implications for the quantization noise performance because the quantization error components are filtered by G_p but not by F_p.

To analyze the effect of the multi rate processing on the noise components we use the following result from [60]: If a W.S.S. random process $e[n]$ with power spectral density $S_e(e^{j\omega})$ is upsampled by a factor of P and then filtered by a filter $G(e^{j\omega})$, as shown in Fig 5.2, then the output process, $q[n]$, is W.S.S. and its power spectral density is given by

$$S_q(e^{j\omega}) = \frac{1}{P} S_e(e^{j\omega P}) |G(e^{j\omega})|^2 \tag{5.3}$$

This is true provided that decimating the impulse response, $g[n]$, by a factor of P does not create any aliasing.

For the system of Fig 5.1, the signal bandwidth of each channel before downsampling and after upsampling is, $\omega_B = \frac{\omega_W}{P} = \frac{\frac{\pi}{RP}}{P} = \frac{\pi}{RP^2}$. So, let us assume that the filters F_p and G_p are ideal band limited "brick wall" filters with individual bandwidths $\frac{\pi}{RP^2}$. As mentioned, the filters $G_p(z)$ have pass band gains of P. Using (5.3), the quantization noise p.s.d. at the output of the filter G_p over its pass-band (it is zero elsewhere) is $S_{ep}(e^{j\omega}) = \frac{1}{P}\frac{\sigma_e^2}{2\pi}P^2 = P\frac{\sigma_e^2}{2\pi}$. The output noise power, σ_{ep}^2 in the signal band of interest is simply the p.s.d. times twice the signal bandwidth. The noise power at the output of each channel is then, $\sigma_{ep}^2 = S_{ep}(e^{j\omega})2\omega_B = \frac{\sigma_e^2}{RP} = \frac{\sigma_e^2}{P}\frac{2f_W}{f_s}$ Assuming that the noise processes at the output of the various channels are all mutually uncorrelated, we can add the in-band noise power of all these channels i.e. multiply the in-band noise power at the output of any one channel by P, to obtain the total in-band noise power as

$$\sigma_{ey}^2 = P\sigma_{ep}^2 = P\frac{\sigma_e^2}{P}\frac{2f_W}{f_s} = \sigma_e^2\frac{2f_W}{f_s} \tag{5.4}$$

Consequently, the P channel sub-band coder PCM converter structure using quantizers operating at f_s achieves a quantization noise performance which is no better than a single PCM converter operating at f_s. Consequently, from the point of view of quantization noise performance, there is *no benefit* in adopting this structure as it consumes P channels of hardware without any improvement in performance.

In the above analysis we assumed that the amplitude distribution at the output of the analysis filters F_p remained essentially unchanged. In practice, if the amplitude range did decrease by, say, a factor of g, the gain in the filters F_p can be increased by a factor of g to make full use of the dynamic range of the internal quantizers. The gain of the synthesis filters G_p can now be reduced by a factor of g and still achieve perfect reconstruction of the signal component. This means that the quantization noise power spectral densities of (5.3) are not amplified as much. In such a scenario, there may be some improvement in SNR performance as the number of channels is increased.

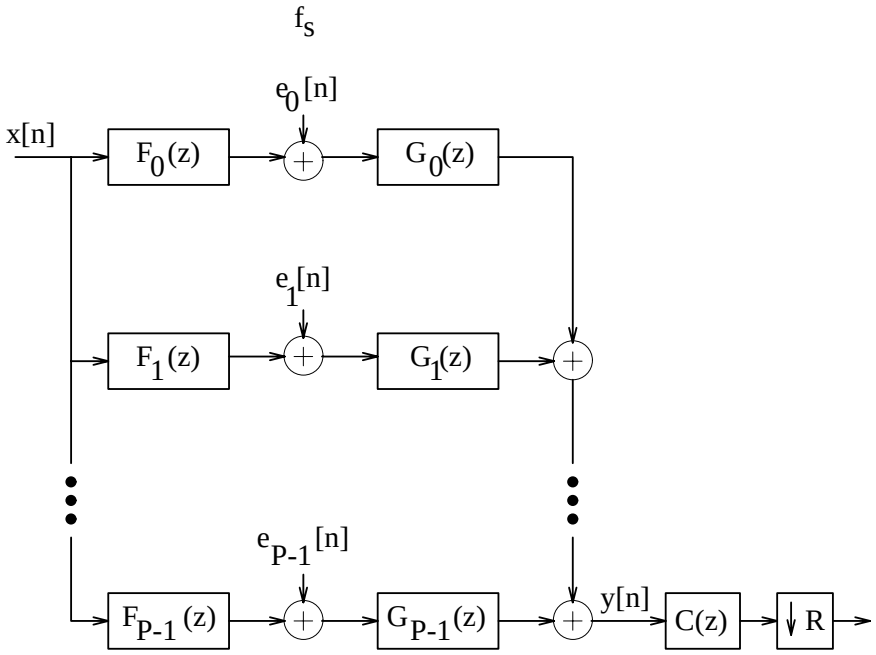


Figure 5.3: Block diagram of a direct multi-band PCM ADC

5.2.2 Frequency Sliced Direct Multi-Band Conversion

Now we describe and analyze a direct multi-band frequency slicing architecture which is motivated by the multi-band delta-sigma structure proposed in Chapter 9. Here we do not perform any upsampling or downsampling as in a sub-band coder structure but simply use P quantizers in parallel operating at f_s. This arrangement is shown in Fig 5.3. Here the output is,

$$Y(z) = \sum_{p=0}^{P-1} X(z)F_p(z)G_p(z) + \sum_{p=0}^{P-1} E_p(z)G_p(z) \tag{5.5}$$

The signal bandwidth of each channel is,

$$\omega_B = \frac{\omega_W}{P} = \frac{\frac{\pi}{R}}{P} \tag{5.6}$$

Let us assume that all $F_p(z) = 1$ (in practice, one would not want to implement long sampled data FIR filters) and that $G_p(z)$ are ideal brickwall filters with bandwidths $\frac{\pi}{RP}$ and passband gain of unity. The p.s.d. of the quantization noise at the output of the filter in the pth channel is,

$$S_{ep}(e^{j\omega}) = \frac{\sigma_e^2}{2\pi}|G_p(e^{j\omega})|^2 \tag{5.7}$$

The noise power at the output of each channel is then,

$$\sigma_{ep}^2 = S_{ep}(e^{j\omega})2\omega_B = \frac{\sigma_e^2}{RP} = \frac{\sigma_e^2}{P}\frac{2f_W}{f_s} \tag{5.8}$$

Assuming that the noise processes at the output of the various channels are all mutually uncorrelated, we can add the in-band noise power of all these channels i.e. multiply the in-band noise power at the output of any one channel by P, to obtain the total in-band noise power as

$$\sigma_{ey}^2 = P\sigma_{ep}^2 = P\frac{\sigma_e^2}{P}\frac{2f_W}{f_s} = \sigma_e^2\frac{2f_W}{f_s} \tag{5.9}$$

The multi-band P channel structure using quantizers operating at f_s achieves a quantization noise performance which is no better than a single PCM A/D operating at f_s. As with the sub-band coder structure, from the point of view of quantization noise performance, there is *no benefit* in adopting this structure as it consumes P channels of hardware without any improvement in performance.

Even if $F_p(z)$ were chosen as brick wall filters with bandwidths $\frac{\pi}{RP}$, the converters performance will not be improved further by increasing P since the maximum amplitude at the output of $F_p(z)$ will not necessarily be reduced.

5.3 Time Sliced PCM Conversion

PCM A/D conversion can be performed by slicing the signal in time by multiplexing or interleaving different time samples to different ADCs. Such a converter, called a time interleaved converter, was proposed in [58]. The signal is sampled at P times the rate at which the ADCs quantize the samples. Each ADC quantizes every Pth sample. Once the ADCs have quantized their respective sample, the samples are demultiplexed back to the higher rate. Fig 5.4 shows a signal processing model of a P channel time interleaved PCM A/D converter. Note that the initial sampling rate is Pf_s, thus the downsamplers, upsamplers, and the delay chains operate at this rate while the quantizers operate at a rate f_s.

Once the quantizers are modeled by the additive noise sources, the entire system is linear. We can now analyze the effect of the system on the signal component and the

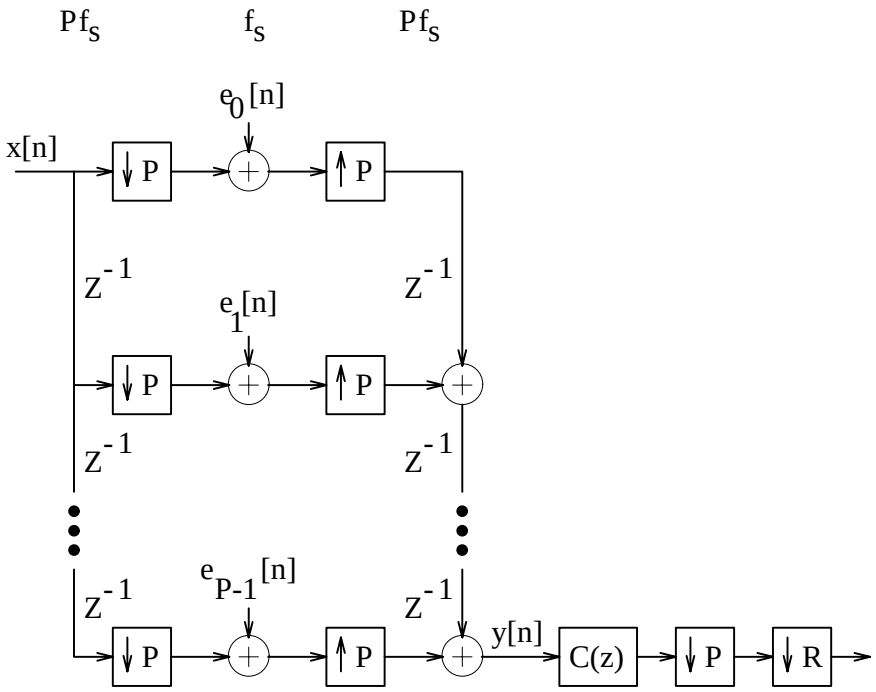


Figure 5.4: Block diagram of a time interleaved PCM ADC

quantization noise components independently. Standard multi rate analysis shows that the signal component experiences perfect reconstruction (reconstruction to within a delay) in passing through the time interleaved system. The output is a delayed version of the input plus an overall quantization noise process, $E(z)$.

$$Y(z) = X(z)z^{-(P-1)} + E(z) \tag{5.10}$$

Again, using multi rate analysis, the overall noise process, $E(z)$ can be written in terms of the quantization noise processes of each quantizer, $E_p(z)$, as

$$E(z) = \sum_{p=0}^{P-1} E_p(z^P)z^{-(P-1-p)} \tag{5.11}$$

The overall noise process $E(z)$ is also a W.S.S. white process with variance σ_e^2. This is proven in Appendix A. Consequently, the power spectral density (p.s.d.) of the output noise process is $S_{ey}(e^{j\omega}) = \frac{\sigma_e^2}{2\pi}$. The total output noise power, σ_{ey}^2 in the signal band of interest is then simply the p.s.d. times twice the total signal bandwidth. The signal

bandwidth is, $\omega_W = \frac{\pi}{RP}$. Consequently,

$$\sigma_{ey}^2 = S_{ey}(e^{j\omega})2\omega_W = \frac{\sigma_e^2}{RP} = \frac{\sigma_e^2}{P}\left(\frac{2f_W}{f_s}\right) \tag{5.12}$$

Note that a single quantizer with variance σ_e^2 operating at rate Pf_s would produce the same in-band output noise power as shown in the above equation (5.12). Consequently, P quantizers operating at a rate f_s are equivalent to one quantizer operating at a rate Pf_s. Moreover, according to (5.12), relative to a single PCM converter operating at a rate f_s, the time interleaved PCM converter experiences a 3 dB improvement in SNR or $\frac{1}{2}$ bit improvement in resolution for every doubling of the number of channels.

5.4 Time Slicing in Terms of Frequency Slicing for Sub-band Coder Structure

We can observe that a time sliced converter is really a sub-band coder with the following analysis and synthesis filters, $F_p(z) = z^{-p}$ and $G_p(z) = z^{-(P-p-1)}$. Plugging these into (5.1) to write the signal component in the output, we obtain

$$Y(z) = \sum_{l=0}^{P-1} X(zW^l)\frac{1}{P}\sum_{p=0}^{P-1} W^{lp} \tag{5.13}$$

However, $\frac{1}{P}\sum_{p=0}^{P-1} W^{lp} = \delta[l]$ and so $Y(z) = X(z)$ to achieve a perfect reconstruction of the signal. Note in this special case, G_p need not have a pass-band gain of P to compensate for the $\frac{1}{P}$ gain since this $\frac{1}{P}$ gain is absorbed in the $\delta[l]$. Also observe that because $G_p(z)$ are all-pass filters, we can not use the quantization analysis of the frequency sliced sub-band coder system. Instead we must use the results of Appendix A.

5.5 PCM Averaging System

We now look at one last system – a PCM averaging system based on the well known idea of averaging different versions of a random process composed of the same signal but uncorrelated noise. Here the input is converted by multiple A/D converters with the same resolution and the outputs of these are averaged. A block diagram is shown in Fig 5.5 The

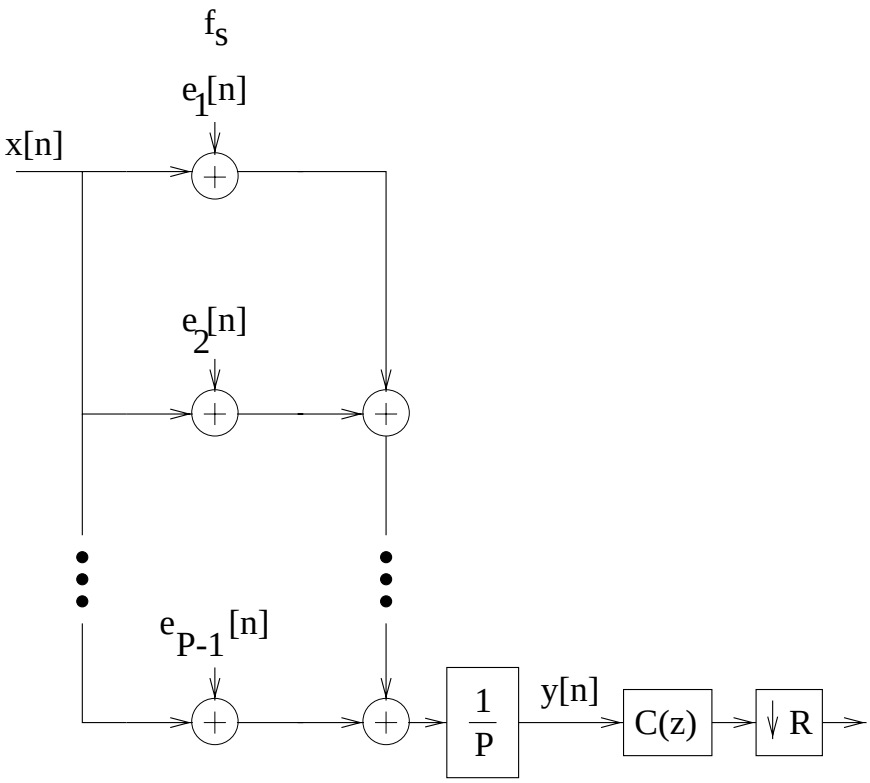


Figure 5.5: Block diagram of a PCM averaging system

output $Y(z)$ can be represented as $Y(z) = X(z) + E(z)$ where

$$E(z) = \frac{1}{P}\sum_{p=0}^{P-1} E_p(z) \tag{5.14}$$

If we assume that the error processes are all mutually uncorrelated (in an actual implementation circuit noise and differences between the circuits of the different converters would make this a reasonable assumption), the quantization noise p.s.d. at output is

$$S_{ey}(e^{j\omega}) = \frac{1}{P} S_e(e^{j\omega}) = \frac{1}{P}\frac{\sigma_e^2}{2\pi} \tag{5.15}$$

The total signal bandwidth is $\omega_W = \frac{\pi}{R}$. Therefore the in-band quantization noise at the output is

$$\sigma_{ey}^2 = S_{ey}(e^{j\omega}) 2\omega_W = \frac{\sigma_e^2}{R} = \frac{\sigma_e^2}{P}\frac{2f_W}{f_s} \tag{5.16}$$

We find that in this case the overall quantization noise performance is indeed the same as that for the time sliced system ! The SNR performance will improve at a rate of 3 dB for every doubling of the number of channels.

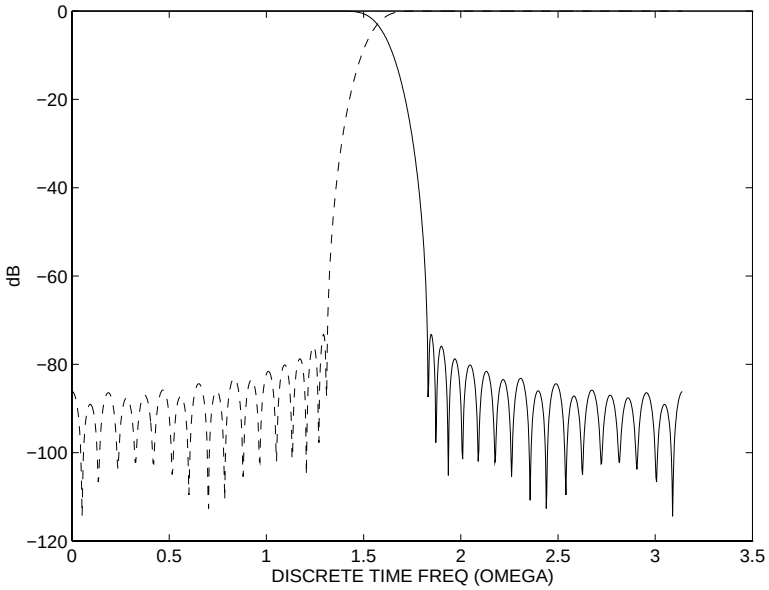


Figure 5.6: Magnitude responses of 2 channel QMF bank analysis filters $F_0(z)$ and $F_1(z)$ (dashed plot). Filter order is 63

5.6 Simulation Results

The conclusions drawn from the above analyses for the frequency and time slice systems have been verified via simulations for the case of doubling the number of channels from 1 to 2 where the overall signal bandwidth is not oversampled i.e. $R = 1$. However, before presenting the results, let us briefly describe the filters used for the sub-band and direct multi-band systems. Although, the analysis was performed assuming brick-wall filters, in practice, we are forced used to use finite length filters. The set of 2 channel perfect reconstruction QMF bank filters provided in [61] has been used to simulate the system of Fig 5.1. The magnitude responses of the analysis filters $F_0(z)$ and $F_1(z)$ are shown in Fig 5.6. In the figure π corresponds with the half the sampling rate at the input of the system. The high-pass filter $F_1(z)$ is obtained from the low-pass filter as $z^{-(M-1)}F_1(z) = F_1(-z^{-1})$ where $M-1$ is the filter order. The synthesis filters are obtained as $G_0(z) = z^{-(M-1)}F_0(z^{-1})$ and $G_1(z) = z^{-(M-1)}F_1(z^{-1})$. For the direct multi-band system of Fig 5.3, $F_0(z) = F_1(z) = 1$ has been chosen. $G_0(z)$ has been designed as a low-pass filter using standard filter design software. We have used a simple filter design procedure in designing $G_1(z)$ such that the system overall response of the filter bank $G_0(z) + G_1(z)$ is perfect reconstruction to within

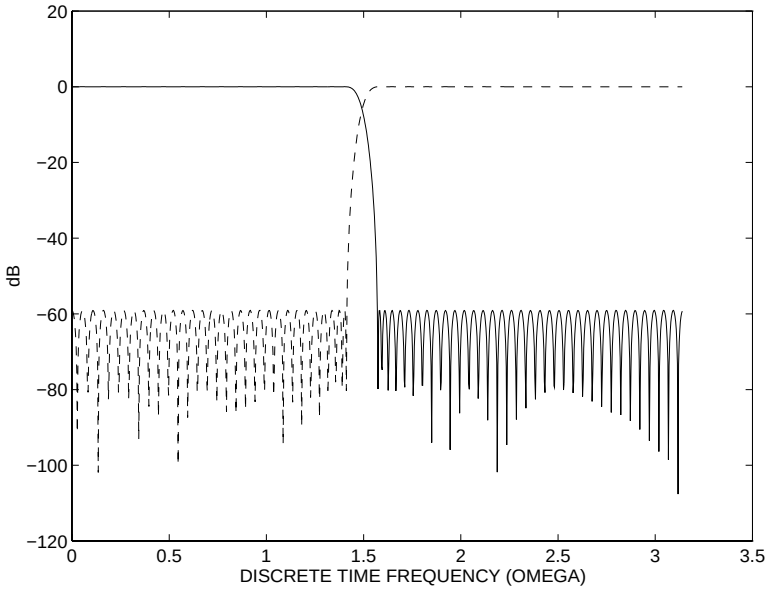


Figure 5.7: Magnitude responses of 2 channel direct multi-band filter bank filters $G_0(z)$ and $G_1(z)$ (dashed plot). Filter order is 128.

a delay. This is done by choosing $G_1(z) = z^{-K} - G_0(z)$ where K is the group delay of $G_0(z)$. It is clear that the overall response of the filter bank is a simple delay. Note that if $G_0(z)$ is a linear phase low-pass filter, $G_1(z)$ is guaranteed to be a high-pass filter with good stop-band attenuation characteristic provided the pass-band ripple of $G_0(z)$ is good. The magnitude responses of the filters $G_0(z)$ and $G_1(z)$ are shown in Fig 5.7. This technique can be easily extended for more than two channels and this is discussed in further detail in Chapter 9.

For the simulations, a uniformly distributed white noise input is used so that the white noise approximation about the quantizer error could be taken for granted. The input is zero mean and uniformly distributed between $[-1, 1]$ so that its variance is $\frac{1}{3}$. The simulations are performed for several values for the number of bits, N, in the quantizer. The reference levels for all the quantizers are chosen as 1 and -1 i.e. they quantize amplitude samples in the range $[-1, 1]$. In all cases, SNR_x is defined to be the SNR which would be obtained by quantizing $x[n]$ at the input to the multi-channel A/D systems at the sampling rate of the input. SNR_y is defined as the SNR which is obtained at the output of the multi-channel A/D systems. It is assumed that for a given set of simulations with a fixed N, all quantizers in the system including the one used to measure SNR_x have the same reference levels of 1

and -1 and same number of bits.

For the convenience of simulations, the same sampling rate is used at the input of all three systems. This is in contrast to the analysis which assumed that all internal quantizers in the three systems operated at the same sampling rate. From our analysis we observe that,

1. For the sub-band system, if the number of channels is doubled from one to two, the SNR will not change when sampling rate for the quantizer in the one channel system is the same as the sampling rate of the quantizers in the two channel system. Equivalently, if the sampling rate of the quantizer in the one channel system is twice that of the quantizers in the two channel system, the SNR of the two channel system will be 3 dB worse than that of the one channel system. Thus, we would expect that in the sub-band system of Fig 5.1, SNR_y will be 3 dB worse than the SNR_x.

2. For the multi-band system, if the number of channels is doubled from one to two, the SNR will not change when sampling rate for the quantizer in the one channel system is the same as the sampling rate of the quantizers in the two channel system. Thus, we would expect that in the multi-band system of Fig 5.3, SNR_x and SNR_y will be the same.

3. For the time interleaved system, if the number of channels is doubled from one to two, the SNR will improve by 3 dB when sampling rate for the quantizer in the one channel system is the same as the sampling rate of the quantizers in the two channel system. Equivalently, if the sampling rate of the quantizer in the one channel system is twice that of the quantizers in the two channel system, the SNR of the two channel system will be the same as the SNR of the one channel system. Thus, we would expect that in the time interleaved system of Fig 5.4, SNR_x and SNR_y will be the same.

These conclusions are verified via simulations whose results are provided in Table 5.1. Since the input to the system is white, the SNR must be measured in the time domain. A comprehensive discussion of SNR measurement techniques can be found in Appendix B.

	sub-band (Fig 5.1)		multi-band (Fig 5.3)		time interleaved (Fig 5.4)	
N	SNR_x	SNR_y	SNR_x	SNR_y	SNR_x	SNR_y
2	12.00	9.03	11.95	11.94	12.05	12.05
4	24.02	21.05	24.19	24.18	24.10	24.10
8	48.16	45.13	48.12	48.13	48.18	48.18
16	96.33	93.30	96.32	96.32	96.31	96.31

Table 5.1: Simulations results for frequency and time sliced PCM systems. N is number of bits in quantizers. SNR_x is the SNR of the input signal quantized with a N bit quantizer. $SNRy$ is the SNR measured at the output of the A/D converter system. All quantizers have the same reference levels of 1 and -1.

5.7 Summary and Comments

The performance of the various converter architectures is summarized in Table 5.2. The table shows the output noise power of a single channel PCM system compared with different P channel PCM architectures with all internal quantizers operating at same rate f_s. Our analsysis was presented in this framework. We observed that the time interleaved and averaging system noise powers are P times less than that of the other systems. Another framework within which to view the performance would be to keep the sampling rate of the input signal the same at a rate of f_s. This means that the internal quantizers of the various systems will now operate at different rates. An equivalent summary of performance viewed in this context is shown in Table 5.3. It was convenient to provide our simulation results in this framework and one can see the correspondence between the relative performance of the SNR_y columns in Table 5.1 and the output noise powers of Table 5.3.

If we examine the performance within the first framework, we observed that the time sliced converter's SNR performance improves by 3 dB for every doubling of the number of channels P, whereas both the sub-band coder and direct multi-band frequency sliced approaches provided no performance improvement if we assumed the the amplitude range of the signal after the analysis filters remained unchanged. Due to the low probability of higher amplitude samples, this amplitude range may decrease slightly for all practical purposes,

slicing	**input rate**	**quantizer rate**	σ_{ey}^2
one channel	f_s	f_s	$\sigma_e^2 \left(\frac{2f_W}{f_s}\right)$
sub-band freq.	Pf_s	f_s	$\sigma_e^2 \left(\frac{2f_W}{f_s}\right)$
direct freq.	f_s	f_s	$\sigma_e^2 \left(\frac{2f_W}{f_s}\right)$
time	Pf_s	f_s	$\frac{\sigma_e^2}{P} \left(\frac{2f_W}{f_s}\right)$
averaging	f_s	f_s	$\frac{\sigma_e^2}{P} \left(\frac{2f_W}{f_s}\right)$

Table 5.2: Output noise power of a single channel PCM system compared with different P channel PCM architectures with all internal quantizers operating at same rate f_s. Signal bandwidth is f_W.

and so the frequency sliced converter SNRs can improve somewhat with every doubling of the number of channels, but certainly not by the full 3 dB improvement achieved by the time sliced converter. For the time sliced and frequency sliced approaches to be equivalent, the outputs of the filters $F_p(z)$ should reduce the amplitude distribution of the input signal by a full factor of $\sqrt{P}$. Note that the **no overload framework** of analysis precluded the system gains from being distributed in such a way as to not adversely affect the quantization noise. Although, the PCM averaging system did perform as well the time sliced system, it is not believed that a true frequency sliced approach exists which can make up for the 3 dB difference with respect to the time sliced converter within the no overload framework.

slicing	**input rate**	**quantizer rate**	σ_{ey}^2
one channel	f_s	f_s	$\sigma_e^2\left(\frac{2f_W}{f_s}\right)$
sub-band freq.	f_s	$\frac{f_s}{P}$	$P\sigma_e^2\left(\frac{2f_W}{f_s}\right)$
direct freq.	f_s	f_s	$\sigma_e^2\left(\frac{2f_W}{f_s}\right)$
time	f_s	$\frac{f_s}{P}$	$\sigma_e^2\left(\frac{2f_W}{f_s}\right)$
averaging	f_s	f_s	$\frac{\sigma_e^2}{P}\left(\frac{2f_W}{f_s}\right)$

Table 5.3: Output noise power of a single channel PCM system compared with different PCM architectures with the signal being sampled at the same rate of f_s in all cases. Signal bandwidth is f_W.

Chapter 6

Complex Band-pass Delta-Sigma Modulation

6.1 Introduction

The discussions on time and frequencly sliced delta-sigma converters entail the use of *complex* band-pass $\Delta\Sigma$ modulators. Therefore, it is appropriate to use this chapter to present what complex $\Delta\Sigma$ modulators are and characterize their performance.

Most band-pass $\Delta\Sigma$ converters [45], [46], [47], use real noise transfer functions (NTFs). By this we mean that all the coefficients in the Z domain transfer function are real. This thesis extends the class of NTFs to allow complex NTFs, which we have described in [62]. The idea of complex NTFs has also been proposed independently in [63]. For a given order NTF, the thesis shows that the use of complex NTFs can improve the resolution that can be obtained for real band-pass signals. We begin by using examples of second order FIR NTFs to intuitively explain why complex NTFs are attractive. Next, we examine and compare the performance of complex and real NTFs in detail for real fourth order FIR, real fourth order IIR, and complex second order FIR NTFs. Finally, we describe a second order complex modulator architecture. This architecture, which is based on transforming a conventional real second order modulator, allows the conversion of a real signal using the complex modulator and recovering the real signal from the complex output of the $\Delta\Sigma$ modulator. We have presented the work described in this chapter in part in [64].

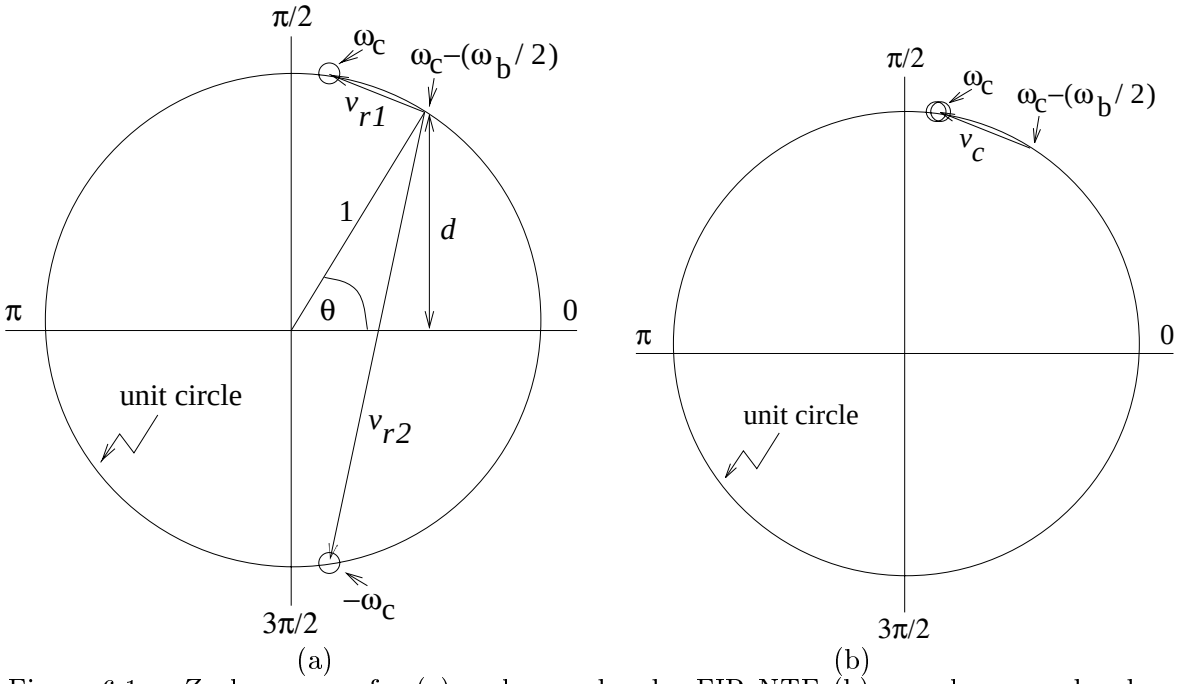


Figure 6.1: Z plane zeros for (a) real second order FIR NTF (b) complex second order FIR NTF

6.2 Intuitive Description of Complex NTF Behavior

Fig 6.1(a) shows the unit circle on the Z plane containing zeros for a real 2nd order FIR NTF. We find it more convenient to use the normalized discrete time frequency $\omega = 2\pi\frac{f}{f_s}$ instead of f in our discussion. A real NTF realizing a zero at center frequency $\omega = \omega_c$ is constrained to also realize one at $-\omega_c$. For a signal bandwidth ω_b, consider the attenuation of the NTF at half the bandwidth away from the zero i.e. at $\omega_c - \frac{\omega_b}{2}$. This attenuation is given by the product of the magnitude of the vectors $\vec{v_{r1}}$ and $\vec{v_{r2}}$ i.e. by $|\vec{v_{r1}}| \times |\vec{v_{r2}}|$. The smaller this product is, the better the attenuation provided by the NTF, and the better the quantization noise suppression over the signal band, and hence the better the SNR which will be obtained. Note that $\vec{v_{r1}}$ represents the contribution from the zero at ω_c while $\vec{v_{r2}}$ represents the contribution from the zero at $-\omega_c$. In the figure, the distance between ω_c and $\omega_c - \frac{\omega_b}{2}$ has been exaggerated − in practice high resolution can be obtained only if the signal is narrowband and this distance is small. As the center frequency ω_c increases, $|\vec{v_{r2}}|$ increases as well and the attenuation provided by the NTF becomes worse and so there will

be less quantization noise suppression. It can be found that

$$|\vec{v_{r1}}| = 2\sin\left(\frac{\omega_b}{4}\right) \tag{6.1}$$

Also for $\omega_b \ll \pi$, $\theta = \omega_c - \frac{\omega_b}{2} \simeq \omega_c$ so that

$$|\vec{v_{r2}}| \simeq 2d = 2\sin(\theta) \simeq 2\sin(\omega_c) \tag{6.2}$$

Note that for $0 < \omega_c < \frac{\pi}{6}, |\vec{v_{r2}}| < 1$ so that the zero at $-\omega_c$ aids the attenuation of the NTF but this aid diminishes with increasing ω_c. For $\omega_c > \frac{\pi}{6}, |\vec{v_{r2}}| > 1$, and the zero at $-\omega_c$ actually decreases the attenuation of the NTF. The performance of real NTFs thus degrades with increasing center frequency, the worst degradation being at $\omega_c = \frac{\pi}{2}$ where $|\vec{v_{r2}}| \simeq 2$ i.e. the NTF attenuation is reduced by 6 dB. This entire behavior is symmetric about $\omega_c = \frac{\pi}{2}$.

For the complex case, the zeros of a second order FIR NTF with both zeros located at a center frequency ω_c, are shown in Fig 6.1(b). The attenuation at $\omega_c - \frac{\omega_b}{2}$ is given by $|\vec{v_c}| \times |\vec{v_c}|$ which depends only on ω_b and not on ω_c as there is no influence from a zero at $-\omega_c$. Consequently, the attenuation provided by the complex NTFs does not suffer any degradation with increasing center frequency. Therefore, at higher center frequencies, complex NTFs can provide more attenuation in the signal band i.e. better quantization noise suppression and so better SNRs than a real transfer function.

6.3 Linearized Analysis Based Performance Comparison of Complex and Real NTFs

A Lth order complex FIR NTF containing L zeros at center frequency ω_c takes the form

$$H_e(z) = \left(1 - z^{-1}e^{j\omega_c}\right)^L \tag{6.3}$$

Using the type of linearized analysis described in Chapter 3, we compute the in-band noise power at the output of this modulator to be,

$$\sigma_n^2 = \frac{4\sigma_e^2\pi^{2L}}{2^{2L+1}(2L+1)}\left(\frac{2f_B}{f_s}\right)^{2L+1} \tag{6.4}$$

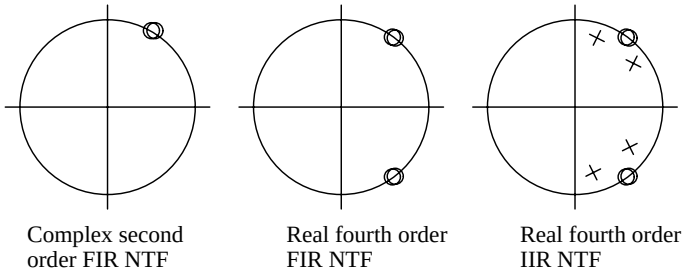


Figure 6.2: Pole / zero configurations

From this we can infer that the SNR performance increases at the rate of $(6L+3)$ dB per octave increment in the oversampling ratio $\frac{f_s}{2f_B}$. This is in contrast to the $(3L+3)$ dB rate of improvement for Lth order real band-pass modulators and the same as the $(6L+3)$ dB rate of improvement for an order $2L$ real band-pass modulator. An order $2L$ real NTF requires $2L$ integrators to implement. However, as one might expect and as we will see later, an order L complex NTF also requires $2L$ integrators to implement. Consequently, it makes sense to compare the performance of order L complex and order $2L$ real transfer functions because, for the same increase in hardware complexity, they result in the same rate of SNR improvement. We will be considering the $L=2$ case in detail i.e. we will be examining complex second order and real fourth order NTFs which assume the use of an 1 bit quantizer.

The second order complex FIR NTF can be described by (6.5)

$$H_e(z) = \left(1 - z^{-1}e^{j\omega_c}\right)^2 = \left(1 - 2z^{-1}e^{j\omega_c} + z^{-2}e^{j2\omega_c}\right) \tag{6.5}$$

The zeros of the transfer function are shown in the first plot Fig 6.2. The fourth order real FIR NTF can be described by (6.6)

$$H_e(z) = \left(1 - z^{-1}e^{j\omega_c}\right)^2 \left(1 - z^{-1}e^{-j\omega_c}\right)^2 \tag{6.6}$$

The zeros of the transfer function are shown in the second plot of Fig 6.2. Although it is well known that placement of NTF zeros distributed over a band of interest can be more effective [65], for ease of comparison we have here assumed coincident zero placement for both the complex and real NTFs.

Two issues affect performance in the absence of implementation imperfections. First, as was explained intuitively, the performance of real FIR band-pass ΔΣ NTFs degrades

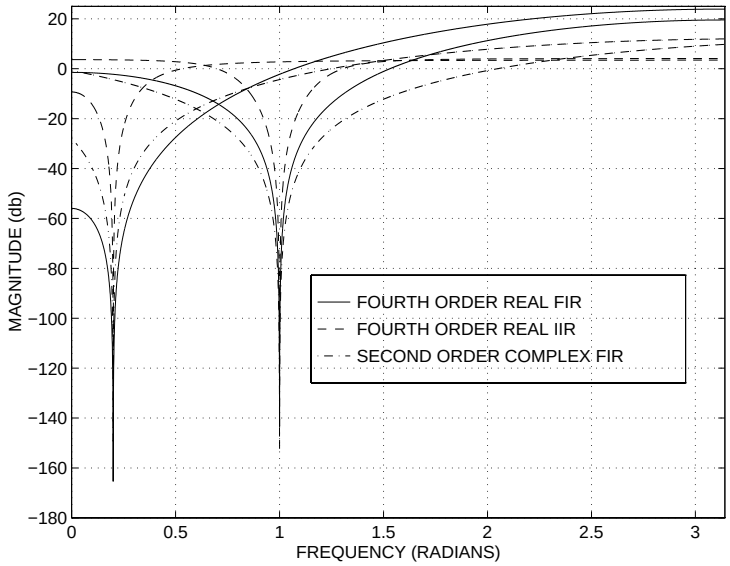


Figure 6.3: Fourth order real NTF magnitude spectra for $R_B = 128$ and center frequencies $\omega_c = 0.2$ and 1 radians. FIR (solid plots) and IIR after pole placement (dashed plots). Also shown for comparison are magnitude spectra of corresponding second order complex FIR NTFs (dashed dot plots) from (6.5).

with increasing center frequency while that of the complex FIR NTF does not change with center frequency. At low center frequencies the fourth order real NTF will produce a higher SNR while at higher center frequencies the second order complex NTF will produce a better SNR (the best being at $\frac{\pi}{2}$). The second issue concerns stability. As is clear from the second part of (6.5), the second order complex NTF is obtained by modulating the coefficients of a standard second order low-pass modulator (whose NTF is $H_e(z) = 1 - 2z^{-1} + z^{-2}$). Consequently, based on the l_1 norm stability criterion [27], a complex second order NTF will be as stable as the low-pass modulator since modulating the NTF coefficients does not change their magnitudes. Higher order FIR NTFs are not unconditionally stable. An ad-hoc stability criterion proposed in [19] is to design the NTF to possess < 6 dB out-of-band gain. As this is in general not achieved by the real fourth order FIR NTFs, the use of poles near the band edges can be used to lower the out-of-band gain to 6 dB or less. The resulting NTF is IIR and a typical pole/zero diagram is shown in the last plot of Fig 6.2.

The introduction of poles, however, degrades the in-band NTF attenuation. The exact degradation of the attenuation depends on the pole radii and angles whose choice may be guided by shaping the STF as well as achieving the stability criterion. We have used a

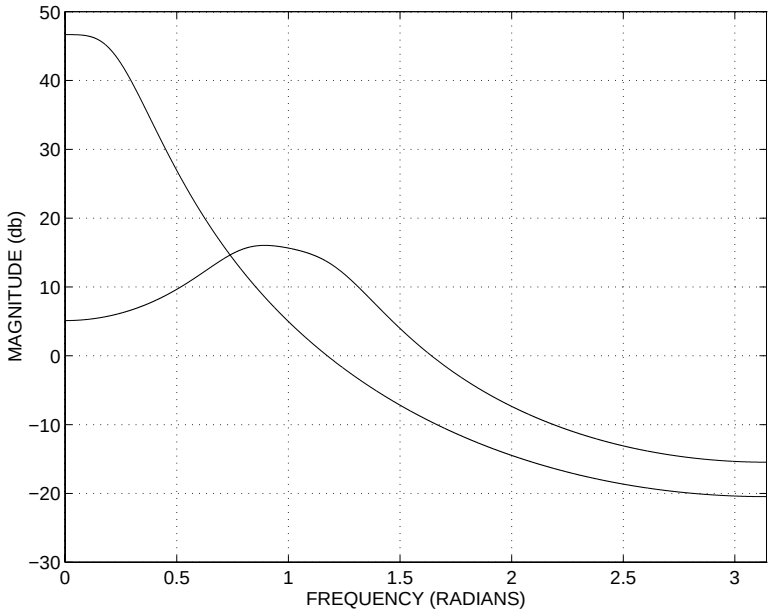


Figure 6.4: Spectra of pole contributions which were used to obtain IIR NTFs from the FIR ones for NTFs in the last figure

least pth optimizer program called filtorX [66], [67] to place the poles such that the out-of-band gain is less than a slightly more conservative value of 4 dB. The NTFs produced by this program are shown in Fig 6.3. The spectra of complex second order FIR NTFs obtained from (6.5) are also shown for comparison. The spectra of the added poles which, converts the FIR NTFs to IIR NTFs to meet the stability criterion, are shown in Fig 6.4. The program optimizes the pole locations to maximize the in-band SNR while meeting the stability criterion. The signal transfer functions produced by the optimizer are band-pass in nature and although they are IIR, they possess approximately linear phase over the signal band. The transfer functions we have considered for comparison are at oversampling ratios of $R_B = 32, 128, 512$. For each R_B, we will consider the following center frequencies $\omega_c = 0.2, 0.6, 1, \frac{\pi}{2}$. Tables 6.1 − 6.3 show the SNR predicted by a linearized analysis of the mentioned NTFs for a full scale input. The NTFs are abbreviated as $cm2$, $rl4_{fir}$, and $rl4$ for the complex second order FIR, real fourth order FIR, and real fourth order IIR NTFs.

We observe that the $cm2$ performs better than the $rl4_{fir}$ NTF only at center frequencies $\omega_c > 0.6$. However, due to the stability constraint, we must actually use the $rl4$ NTF and in this case the SNR degradation due to the presence of poles is enough to make the $cm2$ NTF always outperform the $rl4$ NTF by almost 20 dB for $R_B = 32$ and by 20 dB or more

ω_c	$cm2$	$rl4_{fir}$	$rl4$	$cm2 - rl4_{fir}$	$cm2 - rl4$
0.2 (†)	72.58	91.27	55.04	-18.69	17.54
0.6	72.58	73.32	52.37	-0.74	20.21
1.0	72.58	66.39	52.19	6.19	20.39
$\frac{\pi}{2}$	72.58	63.47	52.60	9.11	19.98

Table 6.1: Oversampling ratio is $R_B = 32$. Comparison of SNRs (dB) obtained by ideal complex second order FIR ($cm2$), real fourth order FIR ($rl4_{fir}$), and real fourth order IIR ($rl4$) transfer functions at different center frequencies, ω_c (radians). (†) For $\omega_c = 0.2$, difficulty in obtaining an optimized NTF with out-of-band gain < 4 dB led to using an optimized NTF with out-of-band gain of about 5 dB.

ω_c	$cm2$	$rl4_{fir}$	$rl4$	$cm2 - rl4_{fir}$	$cm2 - rl4$
0.2	102.68	121.09	76.46	-18.41	26.22
0.6	102.68	102.90	74.64	-0.22	28.04
1.0	102.68	95.95	80.29	6.73	22.39
$\frac{\pi}{2}$	102.68	93.28	83.17	9.40	19.51

Table 6.2: Oversampling ratio is $R_B = 128$. Comparison of SNRs (dB) obtained by ideal complex second order FIR ($cm2$), real fourth order FIR ($rl4_{fir}$), and real fourth order IIR ($rl4$) transfer functions at different center frequencies, ω_c (radians).

ω_c	$cm2$	$rl4_{fir}$	$rl4$	$cm2 - rl4_{fir}$	$cm2 - rl4$
0.2	132.79	151.49	108.31	-18.70	24.48
0.6	132.79	133.35	105.35	-0.56	27.44
1.0	132.79	126.40	105.19	6.39	27.60
$\frac{\pi}{2}$	132.79	123.40	105.17	9.39	27.62

Table 6.3: Oversampling ratio is $R_B = 512$. Comparison of SNRs (dB) obtained by ideal complex second order FIR ($cm2$), real fourth order FIR ($rl4_{fir}$), and real fourth order IIR ($rl4$) transfer functions at different center frequencies, ω_c (radians).

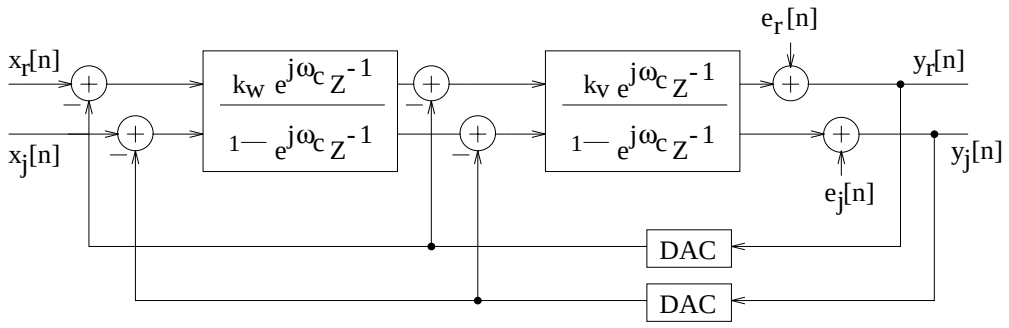


Figure 6.5: Second order complex band-pass delta-sigma modulator

for $R_B = 128$ and 512. Thus, even though the SNR performance of the $cm2$ and $rl4$ NTFs improve at the same rate of about 15 dB per octave increment with the oversampling ratio, the absolute performance of the $cm2$ NTF is always better that of the $rl4$ NTF.

6.4 Complex Second Order Modulator Architecture

Having analyzed the performance of the complex NTF, let us now describe how a $\Delta\Sigma$ modulator architecture can realize the NTF. As indicated earlier, complex NTFs can be obtained by modulating or frequency shifting the transfer function of a real NTF to the center frequency, ω_c, around which the signal band is centered. This has been done by modulating the Z domain NTF and STF coefficients of a low-pass, second order, $\Delta\Sigma$ modulator with the factor $\exp(-j\omega_c) = \exp(-j2\pi\frac{f_c}{f_s})$. The block diagram of the low-pass second order $\Delta\Sigma$ modulator in [7] transformed into a complex band-pass modulator is shown in Fig 6.5. The integrators $\frac{z^{-1}}{1-z^{-1}}$ have been transformed into complex integrators $\frac{z^{-1}e^{j\omega_c}}{1-z^{-1}e^{j\omega_c}}$. With integrator gains $k_w = 0.5$ and $k_v = 2$, the output of the complex second order modulator is given by

$$Y(z) = X(z)z^{-2}\exp(j\omega_c) + [E_r(z) + jE_j(z)][1 - z^{-1}\exp(j\omega_c)]^2 \tag{6.7}$$

where $E(z) = E_r(z) + jE_j(z)$ reflects the fact that there need to be two physical quantizers, E_r and E_j, one for the "real" channel and the other for the "imaginary" channel. The NTF realizes two complex zeros at frequency ω_c. In theory, the input could also be fed into the imaginary channel input of the modulator to produce a STF multiplied by the factor $(1 + j)$ which would imply a gain in signal power of 3 dB. In practice, our simulations have

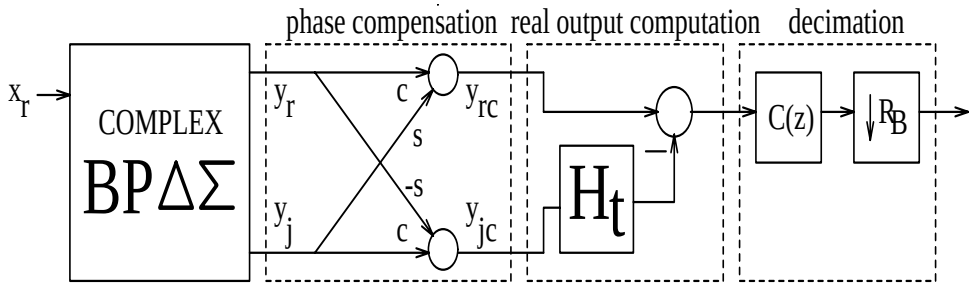


Figure 6.6: Complex BPΔΣ of a real input signal followed by phase compensation, computation of a real output, and decimation. Demodulation of the signal is not shown.

shown this to be not feasible because with the additional input in the imaginary channel, the quantizers overload more often and the increase in quantization overload power does not merit the 3 dB gain in signal power.

Notice that the STF is no longer a pure delay but contains a phase factor. However, this phase factor can be compensated digitally by multiplying the modulator output with the complex constant $\exp(-j\omega_c) = c - js$ where $c = \cos(\frac{2\pi f_c}{f_s})$ and $s = \sin(\frac{2\pi f_c}{f_s})$. The phase compensated output is

$$Y_c(z) = Y(z)\exp(-j\omega_c) \tag{6.8}$$

where $Y_c = Y_{rc} + jY_{jc}$. For a real input signal we have $x_j[n] = 0$ and $X_j(z) = 0$ so that $x[n] = x_r[n]$ and $X(z) = X_r(z)$. In this case,

$$Y_c(z) = X_r(z)z^{-1} + [E_r(z) + jE_j(z)]\exp(-j\omega_c)[1 - z^{-1}\exp(j\omega_c)]^2 \tag{6.9}$$

$Y_c(z)$ consists of a signal term $X_r(z)z^{-1}$, which we have assumed results from a real signal, plus the shaped quantization noise term. From (6.9), we can also see that in the time domain, the signal component in $y_c[n]$ is $x_r[n-1]$. However, $y_c[n]$ will still be complex because of the complex noise and, accordingly, the spectrum of $y_c[n] = y_{rc}[n] + jy_{jc}[n]$ will not be symmetric. A real output can be obtained without disturbing the signal or altering the SNR by considering the final output to be $y_{rc}[n] - y_{jc}[n] * h_t[n]$ where h_t is the impulse response of an ideal Hilbert transformer, $H_t(z)$. In the Z domain, the final output is then $[Y_{rc}(z) - Y_{jc}(z)H_t(z)]$. This operation amounts to keeping only positive frequencies (discarding negative frequencies) with the Hilbert transformer and then taking the real part to make the spectrum symmetric by folding the positive frequencies on to

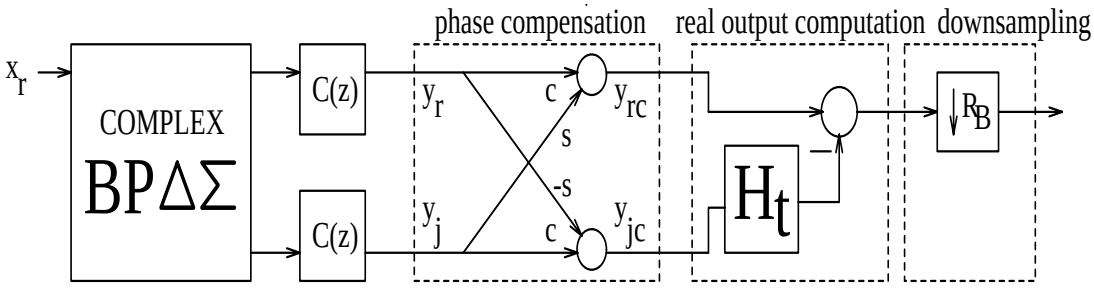


Figure 6.7: Complex BPΔΣ of a real input signal followed by out-of-band noise filtering, phase compensation, computation of a real output, and downsampling. Demodulation of the signal is not shown.

the negative frequency axis. The phase compensation, computation of the real output, and decimation with a decimation filter $C(z)$, are shown in Fig 6.6. Demodulation of the band-pass signal to baseband is not shown here. The order in which the Hilbert transformer operation and decimation are performed affects the digital hardware requirements. For example, Fig 6.7 shows an alternative order in which these operations are performed. In the arrangement of Fig 6.6, the phase compensation operation involves no multiplications since the output of the ΔΣ modulator consists of single bit words but the subsequent Hilbert transformer and decimation filtering operations are performed on multi-bit words. The decimation filter can benefit from reduced computational complexity with a polyphase implementation. In the arrangement of Fig 6.7, the decimation filter $C(z)$ does not benefit from a polyphase implementation but no full hardware multiplications are required for it since its input consists of single bit words from the ΔΣ modulator. The Hilbert transformer in this arrangement can benefit from a reduced complexity polyphase implementation. Note that in practice, it is not possible to design brick-wall decimation filters with reasonable orders and so the downsampling factor may not be R_B but some value D which is slightly less than R_B.

Chapter 7

Practical Performance of Band-pass $\Delta\Sigma$ Modulators

In Chapter 6 we evaluated the theoretical performance of real and complex delta-sigma band-pass systems. Based on behavioral simulations, this chapter provides a quantitative practical performance evaluation of these $\Delta\Sigma$ modulators architectures. This evaluation is performed in detail for modulators with real fourth order and complex second order NTFs. A switched capacitor implementation of the modulators would employ switched capacitor integrators as the building blocks of the modulators. These integrators are in turn realized using operational amplifiers (opamps). In practice these opamps would suffer from finite open loop gain. The NTF coefficients would be realized by scaling signals in different parts of the modulators based on capacitor ratios. In practice, these ratios will not be perfectly matched with the desired ratios. The behavioral simulations model these two major circuit imperfections i.e. finite opamp gain and capacitor ratio mismatch which are found in switched capacitor implementations of the modulators. The modulator structures which allow such behavioral modeling are described. Next, the simulation methodology is briefly described. The simulation results are presented. Simulations are performed for the $cm2$ and $rl4$ NTFs described in Chapter 6. Finally, the feasibility of applying the time sharing concept is investigated for the implementation of the complex modulator.

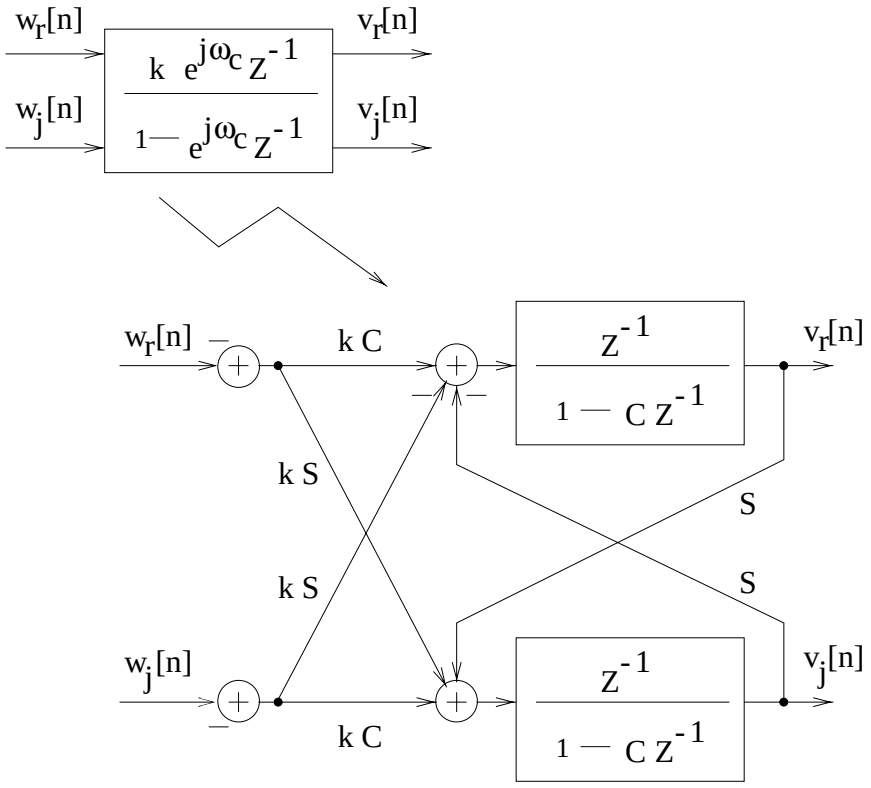


Figure 7.1: Transforming complex integrator block diagram to a physical structure

7.1 Behavioral Modulator Structures

7.1.1 Complex Second Order Modulator

Let us first transform the complex integrator of Fig 6.5 to a physically realizable structure. The diagram of the complex integrator is repeated in the top left hand corner of Fig 7.1 where V denotes the Z domain input and W the Z domain output of the complex integrator. The c and s coefficients are as defined in Chapter 6 i.e. $c = \cos(\omega_c) = \cos(\frac{2\pi f_c}{f_s})$ and $s = \sin(\omega_c) = \sin(\frac{2\pi f_c}{f_s})$. From this complex integrator let us try to extract a structure which reflects what could be physically implemented.

The transfer function of the complex integrator is,

$$\frac{V}{W} = \frac{kz^{-1}e^{j\omega_c}}{1 - z^{-1}e^{j\omega_c}} = \frac{kz^{-1}(c + js)}{1 - z^{-1}(c + js)} \tag{7.1}$$

Writing the complex output V and input W in terms of their real and imaginary components

V_r, W_r and V_j, W_j we obtain

$$\frac{V_r + jV_j}{W_r + jW_j} = \frac{kz^{-1}(c + js)}{1 - z^{-1}(c + js)} \tag{7.2}$$

Cross multiplying and expanding terms we obtain,

$$\begin{aligned} V_r + jV_j - cV_r z^{-1} - jsV_r z^{-1} - jcV_j z^{-1} + sV_j z^{-1} = \\ kcW_r z^{-1} + jksW_r z^{-1} + jkcW_j z^{-1} - ksW_j z^{-1} \end{aligned} \tag{7.3}$$

Grouping terms we have,

$$\begin{aligned} V_r \left(1 - cz^{-1}\right) + jV_j \left(1 - cz^{-1}\right) = \left(kcW_r z^{-1} - ksW_j z^{-1} - sV_j z^{-1}\right) + \\ j\left(kcW_j z^{-1} + ksW_r z^{-1} + sV_r z^{-1}\right) \end{aligned} \tag{7.4}$$

Equating real and imaginary parts of both sides and dividing throughout by $(1 - cz^{-1})$ we obtain,

$$V_r = \frac{kcW_r z^{-1}}{(1 - cz^{-1})} - \frac{ksW_j z^{-1}}{(1 - cz^{-1})} - \frac{sV_j z^{-1}}{(1 - cz^{-1})} \tag{7.5}$$

$$V_j = \frac{kcW_j z^{-1}}{(1 - cz^{-1})} + \frac{ksW_r z^{-1}}{(1 - cz^{-1})} + \frac{sV_r z^{-1}}{(1 - cz^{-1})} \tag{7.6}$$

We recognize that the real output, V_r, is composed of a weighted sum of three signals modified by a transfer function. The three signals are: the real input, W_r, the imaginary input, W_j, and the imaginary output, V_j. The common transfer function which these three signals experience is $\frac{z^{-1}}{1-cz^{-1}}$ which is the transfer function of a damped integrator with dampling factor c. Likewise, the imaginary output, V_j, is composed of a weighted sum of three signals modified by a transfer function. In this case, the three signals are: the imaginary input, W_j, the real input, W_r, and the real output, V_r. The common transfer function experienced by the signals which form V_j is again the damped integrator $\frac{z^{-1}}{1-cz^{-1}}$. If we examine the above equations we can observe that the input cross couplings give rise to the complex factor in the numerator of the complex integrator's transfer function, $\frac{e^{j\omega_c} z^{-1}}{1-e^{j\omega_c} z^{-1}}$, while the output cross couplings give rise to the complex factor in the denominator of the complex integrator's transfer function. Fig 7.2 shows a behavioral model of the complex second order modulator of Fig 6.5 which reflects the transformation of the complex integrator to a physically realizable form based on (7.5) and (7.6).

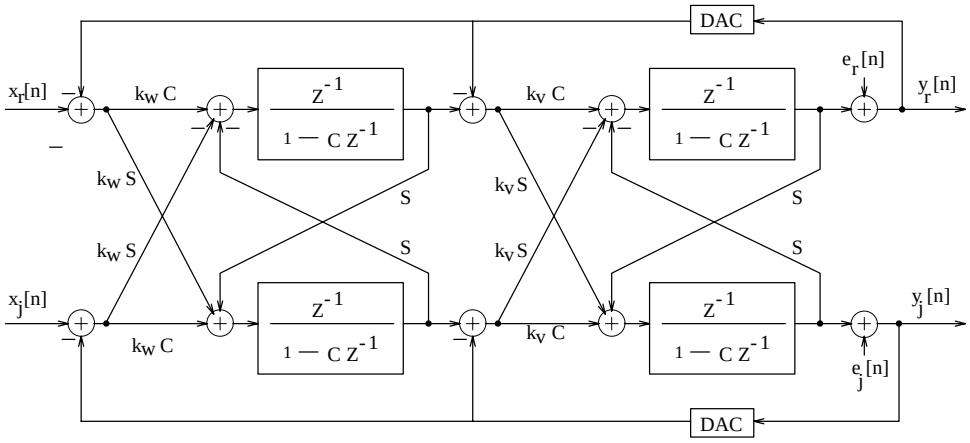


Figure 7.2: Behavioral model of complex 2nd order band-pass modulator

The simulations of the modulator are performed in the time domain by simulating the difference equation for the structure. The time domain version of (7.5) and (7.6) are:

$$v_r[n] = (kcw_r[n-1] - ksw_j[n-1] - sv_j[n-1]) + cv_r[n-1] \tag{7.7}$$

$$v_j[n] = (kcw_j[n-1] + ksw_r[n-1] + sv_r[n-1]) + cv_j[n-1] \tag{7.8}$$

In a switched capacitor implementation, the kc, ks, s weighting values as well as the dampling factor c will be realized as capacitor ratios and there will be one capacitor ratio for each term. Consequently, capacitor ratio mismatches with respect to their ideal values are modeled as non-ideal coefficients. Finite gain of the opamp is modeled by modifying the difference equation and hence transfer function of the damped integrator. In general, let $w_k[n], k = 0 \ldots K-1$ be various inputs which are each weighted by capacitor ratios g_k before being summed as the overall input to the damped integrator. If $v[n]$ is the output of the integrator then, in the ideal situation, we have

$$v[n] = \left(\sum_{k=0}^{K-1} g_k w_k[n-1] \right) + c\, v[n-1] \tag{7.9}$$

An analysis of a typical switched capacitor implementation of the behavioral model reveals that with a finite opamp gain of A, the modified difference equation is now,

$$v[n] = g \left(\sum_{k=0}^{K-1} g_k w_k[n-1] \right) + \alpha\, c\, v[n-1] \tag{7.10}$$

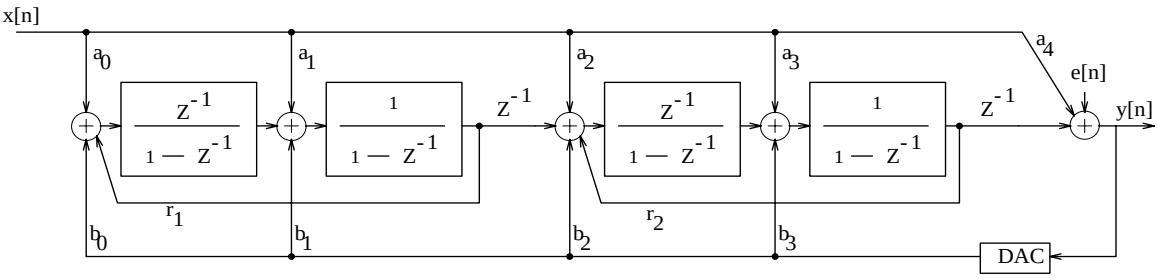


Figure 7.3: Behavioral model of 4th order cascade of resonator band-pass modulator

where g and α are

$$g = \frac{1}{1 + \frac{1+\sum_{k=0}^{K-1} g_k}{A}} \tag{7.11}$$

$$\alpha = \frac{1}{1 + \frac{\sum_{k=0}^{K-1} g_k}{1+A}} \tag{7.12}$$

Consequently, we can simulate the performance of the modulator with imprecise capacitor ratios and finite opamp gain by using the modified difference equation (7.10) for the integrators. This method of modeling the imperfections is applied to (7.7) and (7.8).

7.1.2 Real Fourth Order Modulator

A behavioral model is also used to simulate the SNR performance of the real fourth order modulator in the presence of imperfections such as capacitor ratio mismatch and finite opamp gain. The modulator structure used is the cascade of resonators which has been examined and found to be a good candidate for band-pass $\Delta\Sigma$ modulators [20], [46]. Fig 7.3 shows a behavioral model of this structure. The various a, b, and r coefficients shown in the structure are calculated from the coefficients of the signal and noise transfer functions designed in Chapter 6. In a switched capacitor implementation, the a, b, and r coefficients would be realized as capacitor ratios. Therefore, capacitor ratio mismatches are modeled by using non-ideal values for these coefficients.

Finite opamp gain is modeled in a way similar to that described for the complex second order modulator. The difference here is that a lossless integrator is used. Its ideal transfer function is $\frac{z^{-1}}{1-z^{-1}}$. Let $w_k[n], k = 0 \ldots K-1$ be various inputs which are each weighted by capacitor ratios g_k before being summed as the overall input to the integrator. If $v[n]$ is the

output of the integrator then, in the ideal situation, we have,

$$v[n] = \left(\sum_{k=0}^{K-1} g_k w_k[n-1] \right) + v[n-1] \tag{7.13}$$

An analysis of a typical switched capacitor implementation of the behavioral model reveals that with a finite opamp gain of A, the modified difference equation is now,

$$v[n] = g \left(\sum_{k=0}^{K-1} g_k w_k[n-1] \right) + \alpha\, v[n-1] \tag{7.14}$$

where g and α are as defined in (7.11) and (7.12) respectively.

7.2 Simulation Study

7.2.1 Simulation Conditions and Parameters

Simulations were performed with the real fourth order and complex second order modulator using a single tone input. The tone amplitude was chosen as 0.2 (-17 dB) relative to the quantization reference level. The tone frequency was chosen to prevent spectral leakage. Since the focus here is on the performance of the modulators with respect to analog circuit imperfections, the in-band SNRs were measured at the output of the modulators assuming ideal digital filters i.e. no explicit digital filtering is performed. This was possible since the SNRs were measured in the frequency domain and could be measured only over the signal band of interest. Also, for the second order complex modulator, an ideal digital Hilbert transformer filter is assumed. For the oversampling ratios up to $R_B = 128$, 64K point FFTs were used in the SNR computation while for $R_B = 512$, a 256K point FFT was used. A Blackman window was used for all the SNR measurments. A comprehensive description of SNR calculation techniques can be found in Appendix B.

Simulations have been performed for the transfer functions designed in Chapter 6. Therefore, each modulator is simulated at oversampling ratios of $R_B = 32, 128$, and 512 for the transfer functions centered at $\omega_c = 0.2, 0.6, 1, \frac{\pi}{2}$ radians. For each set of (R_B, ω_c), effects of finite opamp open loop gain and capacitor ratio mismatch were examined separately. Finite opamp gains are varied between 40 dB and 80 dB in increments of 10 dB assuming ideal capacitor ratios. For each of set of (R_B, ω_c), all the capacitor ratios in the modulator

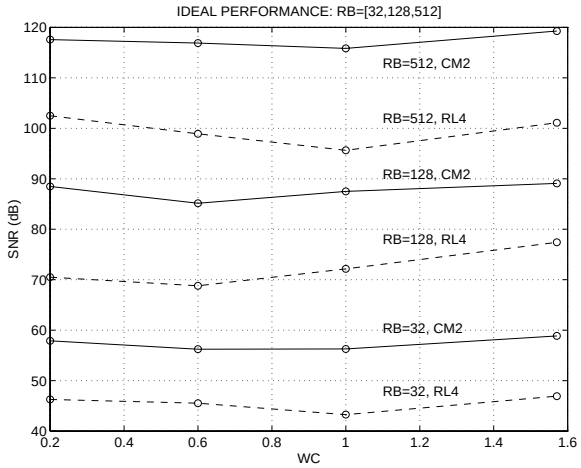
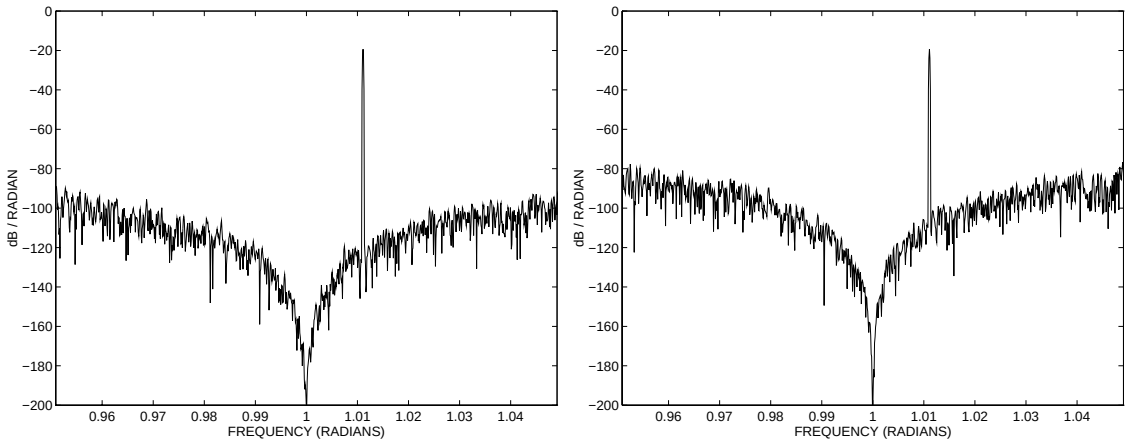


Figure 7.4: Simulation SNRs (dB) vs ω_c = center frequency (radians) for complex second order (solid line) and real fourth order (dashed line) modulators. Each curve curve corresponds to a different oversampling ratio of $R_B = 32$, 128, 512. Ideal (infinite) opamp gains for integrators and ideal (no mismatch) capacitor ratios are assumed.

structures are varied randomly assuming ideal opamp gains. Each capacitor ratio is varied by either 0.1, 0.2, 0.5 and 1% from its ideal value. For each mismatch percent, Monte Carlo simulations are performed with 200 simulations. For several test cases, it was verified that performing 400 and 1000 simulations did not yield significantly more information and so 200 simulations were considered sufficient. These simulation results are discussed here.

7.2.2 Simulation Results

We now examine the SNR performance of the complex second order and real fourth order modulators with an ideal as well as imperfect implementation. Fig 7.4 shows the simulation SNR obtained with an ideal implementation for the two modulators using a tone input. Fig 7.5 shows typical power spectrum outputs for the complex second order and real fourth order modulators. Unlike the linearized results of Tables 6.1-6.3 which assume a full scale tone input with -3 dB signal power, the signal power of the tones used for simulation is -17 dB. The results obtained are shown in Fig 7.4. Compensating for this difference in signal power, we observe that the linearized SNRs and simulation SNRs for the $cm2$ modulator are quite close. For the $rl4$ modulator, the simulation SNR is consistently higher than the linearized model SNR even though the in-band shaped noise observed in the power spectrum of the modulator output is still relatively white. This difference between the



(a) (b)

Figure 7.5: Power spectra output of (a) complex second order modulator (b) real fourth order modulator. $R_B = 128$, $\omega_c = 1$, spectra detail shown over region $[\omega_c - 2\omega_B, \omega_c + 2\omega_B]$.

linearized analysis and the actual simulation SNRs for the real fourth order system, works in its favor – for example, based on the simulation SNRs, the $cm2$ modulator performs only 12 dB (instead of 18-20 dB) or so better than the $rl4$ at $R = 32$ while at the higher oversampling ratios, the performance improvement is similarly reduced by up to 6-8 dB.

We now explore the degradation in SNR with implementation imperfections. This is done for the oversampling ratios and center frequencies of Fig 7.4. The degradation is always taken to be relative to the ideal simulation SNR data shown in Fig 7.4.

Complex Second Order Modulator

Finite opamp gain effects

As can be expected, the SNR performance degrades with finite opamp gain. Fig 7.6 shows how the SNR degrades as a function of opamp gain for various values of $R_B = 32, 128, 512$ and for center frequencies $\omega_c = 0.2, 0.6, 1.0$ and $\frac{\pi}{2}$. As with the standard low-pass second order modulator, the SNR degradation is greater for larger oversampling ratios. The degradation increases somewhat with increasing center frequency. In all cases, with 65 dB opamp open loop gain, the SNR loss is less than 6 dB. The above results assumed that all four opamps, needed to implement the four integrators of Fig 7.2, possessed the same albeit non-ideal gains. Simulations were also performed to examine the performance when the

opamp gains were not matched. Figs 7.7(a)-(d) show the effects of opamp gain mismatch for $R_B = 32, 128$ at $\omega_c = 0.2, 1.0$. Each diagram plots SNR degradation versus the opamp gain mismatch for different nominal values of the opamp gain i.e. each curve is for a different nominal gain. The larger the nominal gain, the less severe the matching problems. For a given nominal opamp gain, the SNR degradation is worse at larger oversampling ratios. For a given nominal opamp gain and given R_B, in general, the SNR degradation becomes somewhat worse at higher center frequencies. We observe that with a 60 dB opamp gain and with 10% mismatch among the gains, the SNR degradation is already 5 dB for $R_B = 32$ and $\omega_c = 1.0$. The degradation increases to about 28 dB for $R_B = 128$. Since the matching of opamp open loop gains is difficult the performance of the complex modulator is seriously affected by finite opamp gain.

Capacitor ratio mismatch effects

We now explore the effect of mismatch among the various capacitor ratios in the modulator structure of Fig 7.2. Monte Carlo simulations were performed by varying all the ratios randomly by a certain percent from their nominal values. For each pair of (R_B, ω_c,percent mismatch) 200 simulations were run. In Fig 7.8 the worst case SNR degradation is plotted versus percent mismatch for 0.1, 0.2, 0.5 and 1 percent ratio mismatch. Each plot has three different curves corresponding with different values of $R_B = 32, 128, 512$. The sensitivity to capacitor ratio mismatch clearly increases with oversampling ratio. It also increases with center frequency. The performance is very sensitive to capacitor ratio mismatch. Even with 0.1% mismatch, the worst SNR degradation at $R_B = 128$ is more than 30 dB at $\omega_c = 1.0$.

Real Fourth Order Modulator

Finite opamp gain effects

The real fourth order modulator, of course, will also suffer from performance degradation due to finite opamp open loop gain. Fig 7.9 plots SNR degradation vs opamp gain at $\omega_c = 0.2$ and 1.0 for $R_B = 32, 128$, and 512. Although there is some sensitivity to opamp gain, with 65 dB opamp gain there is only 5 dB degradation for $R_B = 512$ and $\omega_c = \frac{\pi}{2}$. This is true if all the opamps used to implement the four integratiors of Fig 7.3 are the same. Fig 7.10 shows the effect of opamp gain mismatch. The SNR loss is plotted versus

percent mismatch in the opamp gains. Each plot contains several curves corresponding to different values of the nominal gains. Here we find that, as for the complex modulator, the SNR penalty is worse for larger overampling ratios and somewhat worse at higher center frequencies. However, the performance degradation is significantly less. Even with a nominal gain of 60 dB, there is less than 5 dB SNR drop for $R_B = 128$ and $\omega_c = 1.0$. Moreover, as is clear from the relatively flat curves of Fig 7.10, the drop is largely insensitive to the degree of matching among the opamp gains.

Capacitor ratio mismatch effects

We also explored the effect of mismatch among the various capacitor ratios of the modulator structure of Fig 7.3. Again, Monte Carlo simulations were performed by varying all the ratios randomly by a certain percent from their nominal values. For each pair of (R_B, ω_c, percent mismatch) 200 simulations were run. In Fig 7.11 the worst case SNR degradation is plotted versus percent mismatch for 0.1, 0.2, 0.5 and 1 percent ratio mismatch. Each plot has three different curves corresponding with different values of $R_B = 32, 128, 512$. As before, the sensitivity to capacitor ratio mismatch increases with oversampling ratio and with center frequency. However, the performance is much less sensitive to mismatch than that of the complex modulator. For example, at $\omega_c = 1.0$ and $R_B = 128$, the worse SNR degradation is only 6 dB with 0.5% mismatch and is about a couple of 2 dB with 0.1% mismatch.

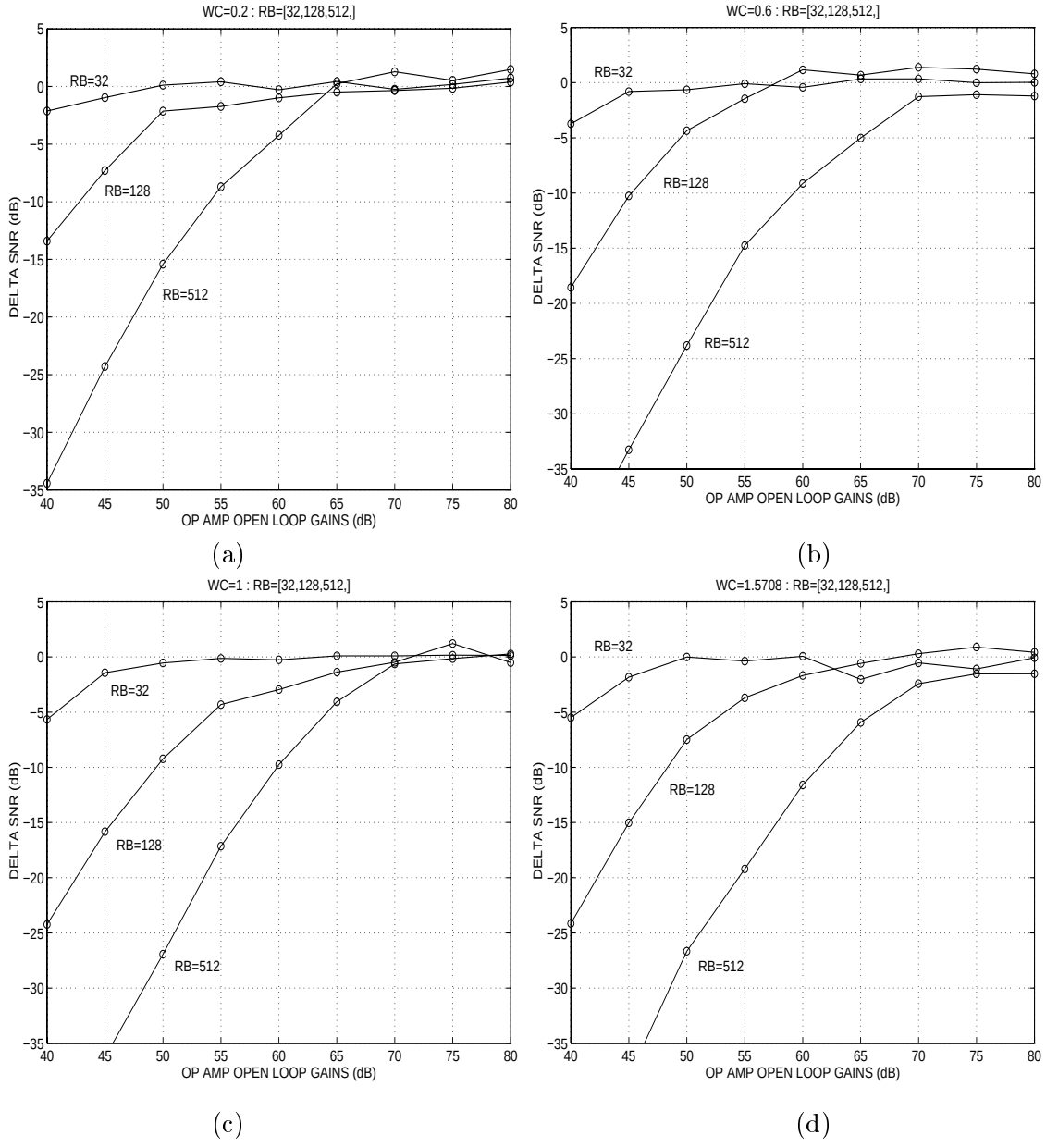


Figure 7.6: Complex second order modulator: ΔSNR (dB) vs opamp open loop gain (dB) for (a) $\omega_c = 0.2$ (b) $\omega_c = 0.6$ (c) $\omega_c = 1$ (d) $\omega_c = \frac{\pi}{2}$. The curves in each plot correspond to different oversampling ratios of $R_B = 32$, 128, 512.

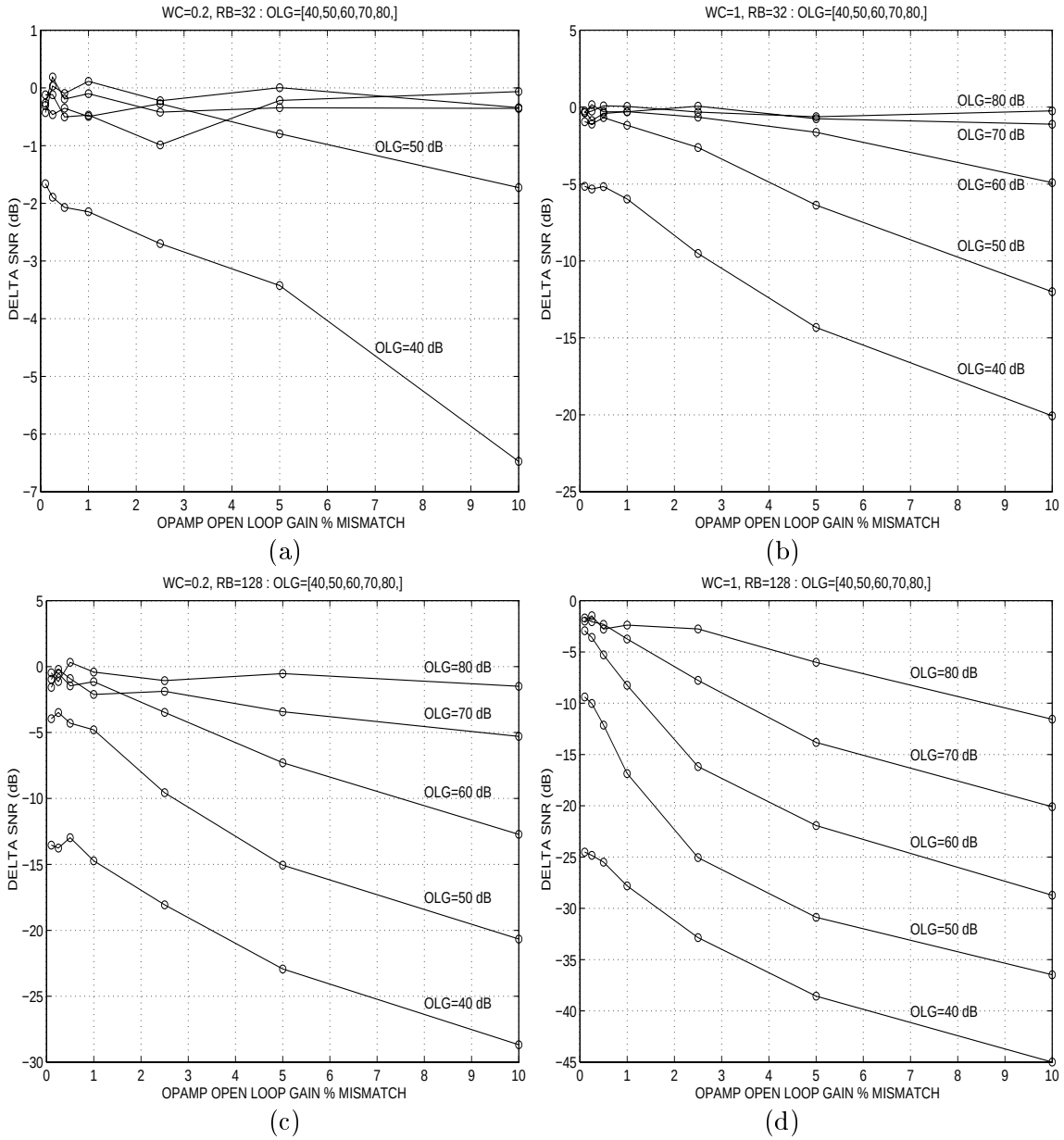


Figure 7.7: Complex second order modulator: ΔSNR (dB) vs opamp open loop gain mismatch percent for (a) R_B=32, $\omega_c = 0.2$ (b) $R_B = 32$, $\omega_c = 1.0$ (c) R_B=128, $\omega_c = 0.2$ (d) $R_B = 128$, $\omega_c = 1.0$. Each plot has five different curves corresponding to different nominal open loop gains (OLG) of 40, 50, 60, 70, and 80 dB.

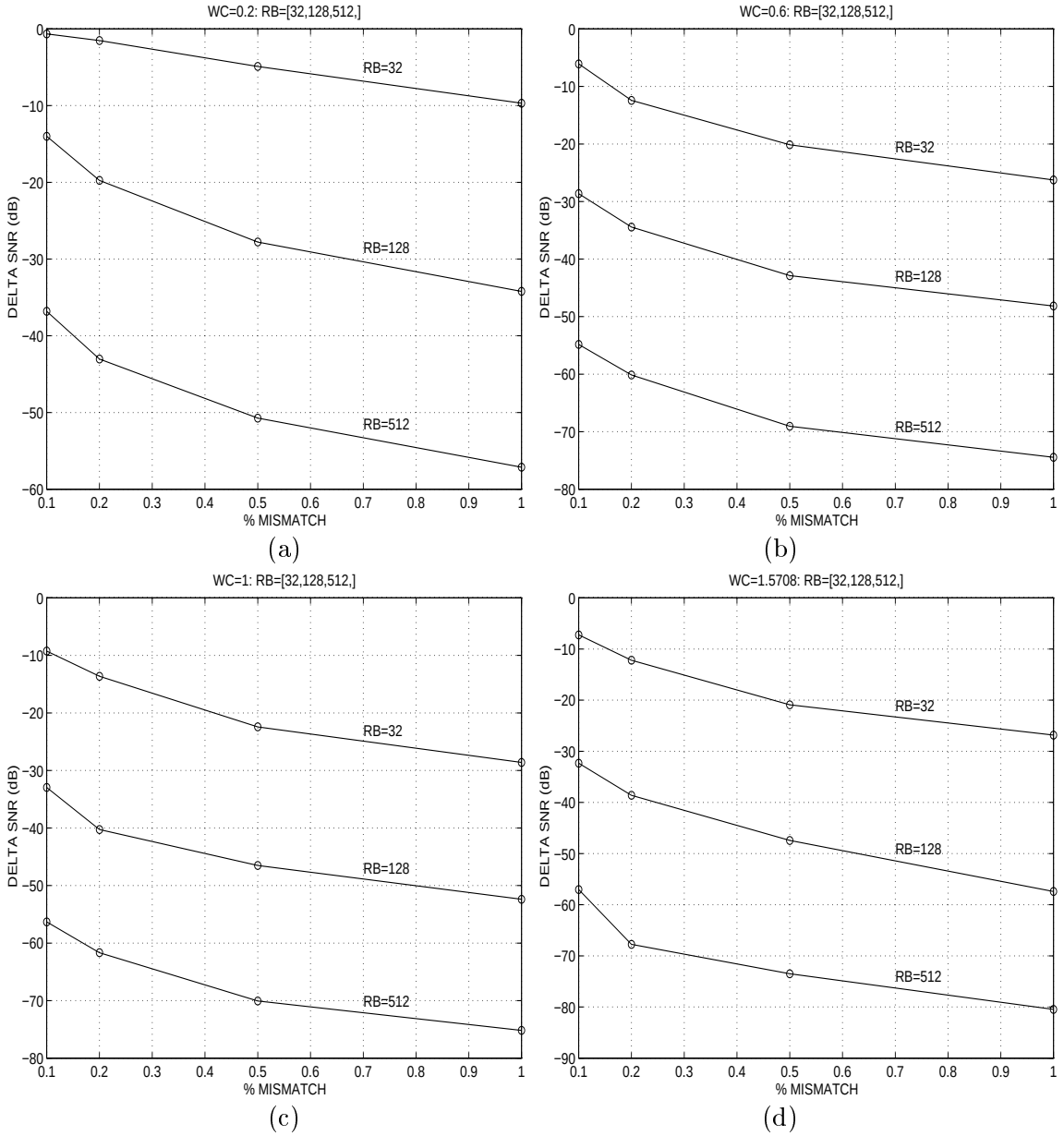


Figure 7.8: Complex second order modulator: worst case ΔSNR (dB) vs percent capacitor ratio mismatch using Monte carlo simulations with 200 iterations. (a) $\omega_c = 0.2$ (b) $\omega_c = 0.6$ (c) $\omega_c = 1.0$ (d) $\omega_c = \frac{\pi}{2}$. Each plot has three curves corresponding to oversampling ratios of $R_B = 32, 128, 512$.

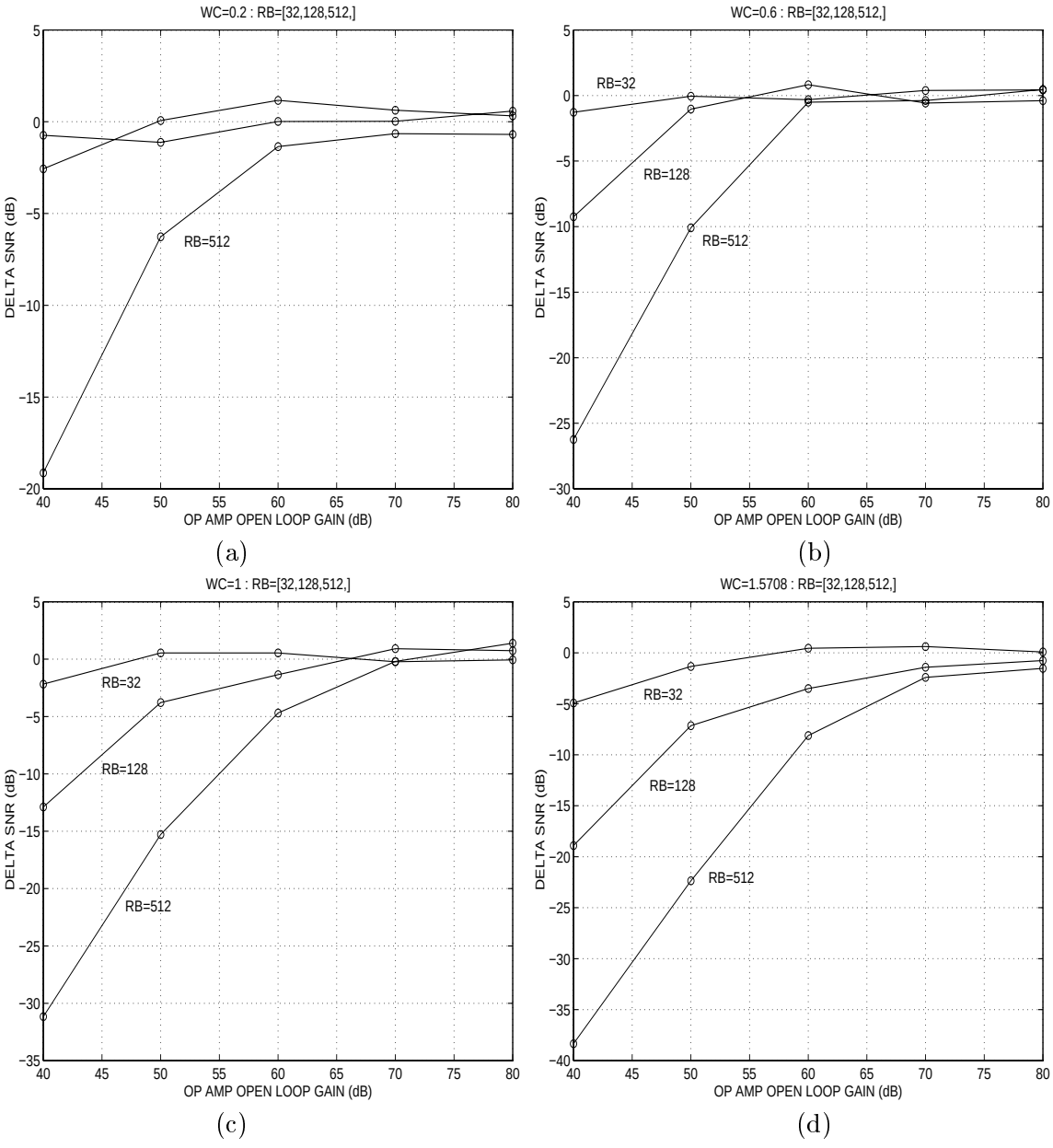


Figure 7.9: Real fourth order modulator: ΔSNR (dB) vs opamp open loop gain (dB) for (a) $\omega_c = 0.2$ (b) $\omega_c = 0.6$ (c) $\omega_c = 1$ (d) $\omega_c = \frac{\pi}{2}$. Each plot has different curves corresponding to oversampling ratios of $R_B = 32$, 128, and 512.

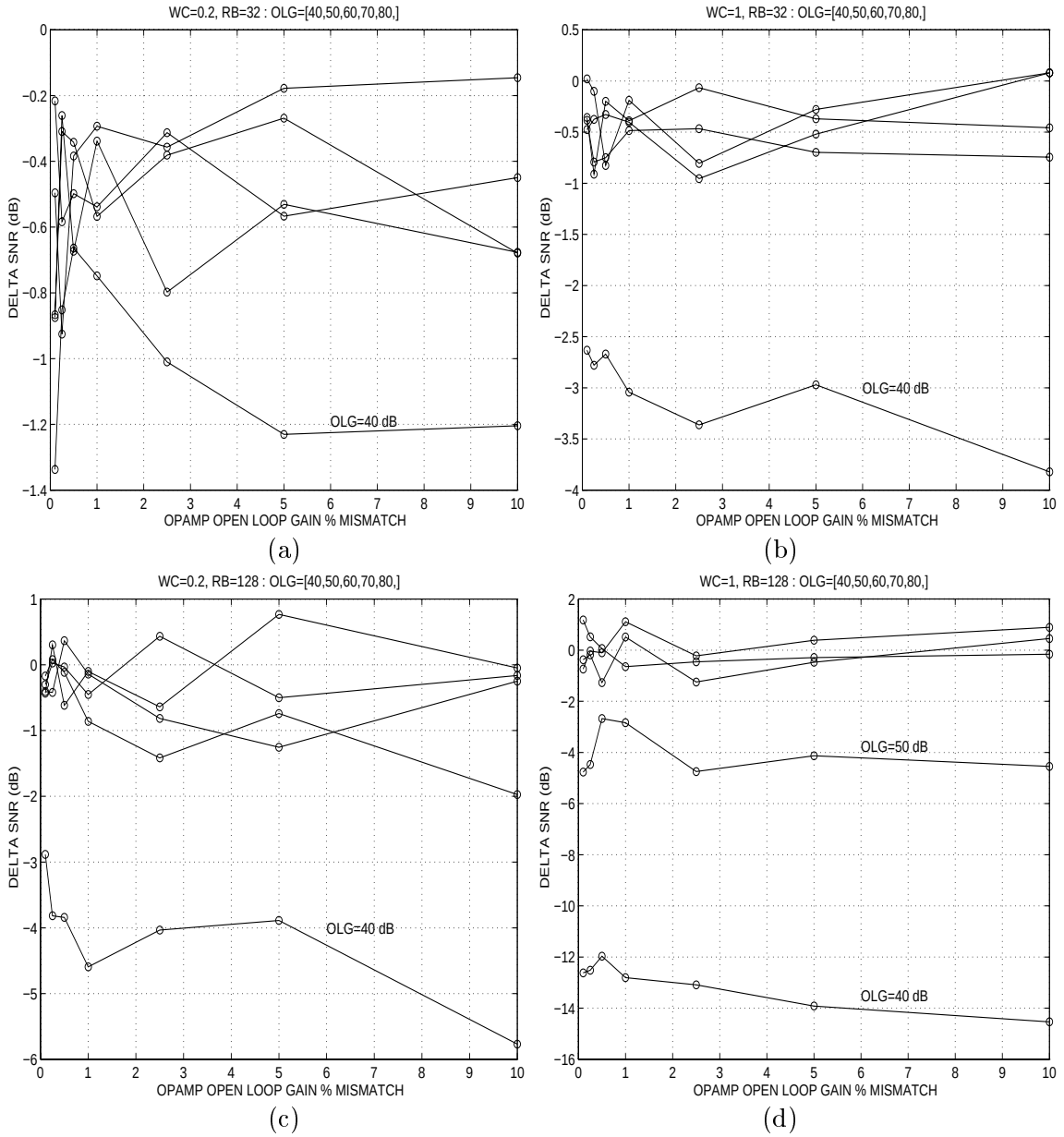


Figure 7.10: Real fourth order modulator: ΔSNR vs opamp open loop gain mismatch for (a) R_B=32, $\omega_c = 0.2$ (b) $R_B = 32$, $\omega_c = 1.0$ (c) R_B=128, $\omega_c = 0.2$ (d) $R_B = 128$, $\omega_c = 1.0$. Each plot has five different curves corresponding to nominal open loop gains (OLG) of 40, 50, 60, 70, and 80 dB.

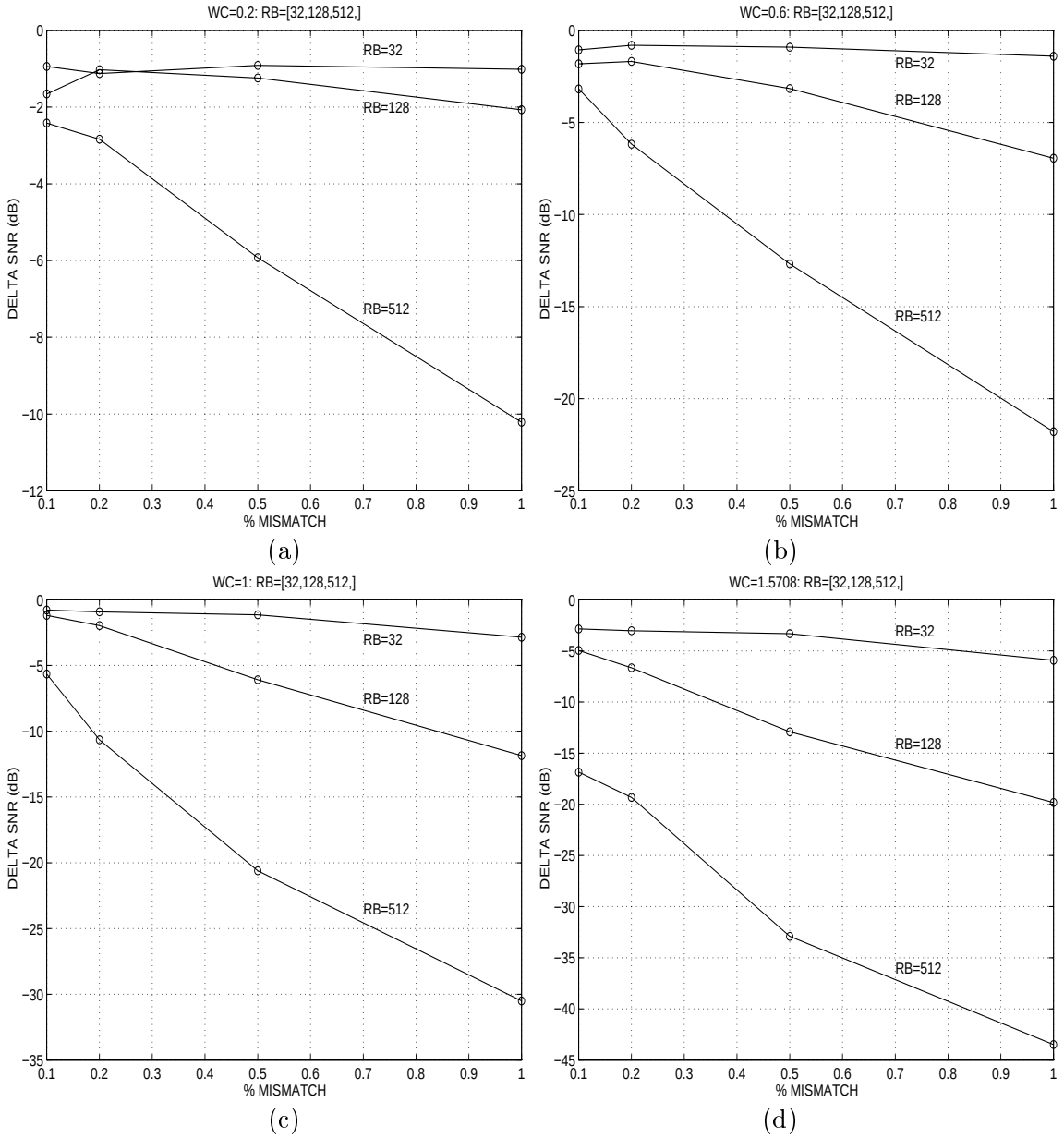


Figure 7.11: Real fourth order modulator: worst case ΔSNR vs percent capacitor ratio mismatch using Monte carlo simulations with 200 iterations. (a) $\omega_c = 0.2$ (b) $\omega_c = 0.6$ (c) $\omega_c = 1.0$ (d) $\omega_c = \frac{\pi}{2}$. Each plot has three curves corresponding to oversampling ratios of $R_B = 32, 128, 512$.

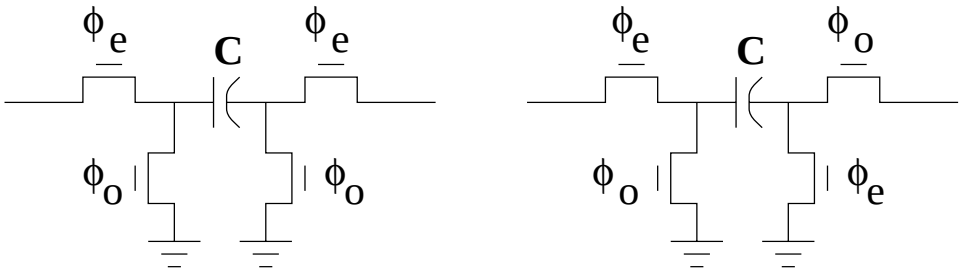


Figure 7.12: Parasitic insensitive switched capacitor configurations

7.3 Applying Time Sharing to The Complex Integrator

The simulation results presented show that the complex integrator performance suffers greatly from mismatch in the opamp gain between the real and imaginary channels when the opamp gain is not infinite. Mismatch among capacitor ratios also presents a significant performance degradation. Although techniques such as described in [69], [70] could possibly be applied to achieve higher opamp gains, extremely large gains are still required to eliminate mismatch effects between real and imaginary channels especially for high oversampling ratios. In this section we investigate the feasibility of applying the time sharing concept [71], [72] to the complex integrator. Time shared circuits realize two different transfer functions using the same opamp. This is possible if the switch configurations can be chosen such that such that the opamp is effectively used for each transfer function on one phase of the two possible phases in a two phased clock system. Consequently, there exists the potential to eliminate performance degradation due to opamp gain mismatch if the real and imagninary channels of the complex integrator can operate with the same opamp.

7.3.1 Basic Complex Switched Capacitor Integrator Circuit

Switched capacitor implementations of complex filters have been considered in [73] and extended to complex $\Delta\Sigma$ modulators in [63]. Let us now disscuss in further detail the fundamentals of the implementation of the switched capacitor integrators before proceeding to our discussion on time sharing. Fig 7.12 shows the two parasitic insensitive switched capacitors which are used as building blocks in many switched capacitor circuits. From now on we will denote the configuration of Fig 7.12(a) by the acronym EE to reflect the

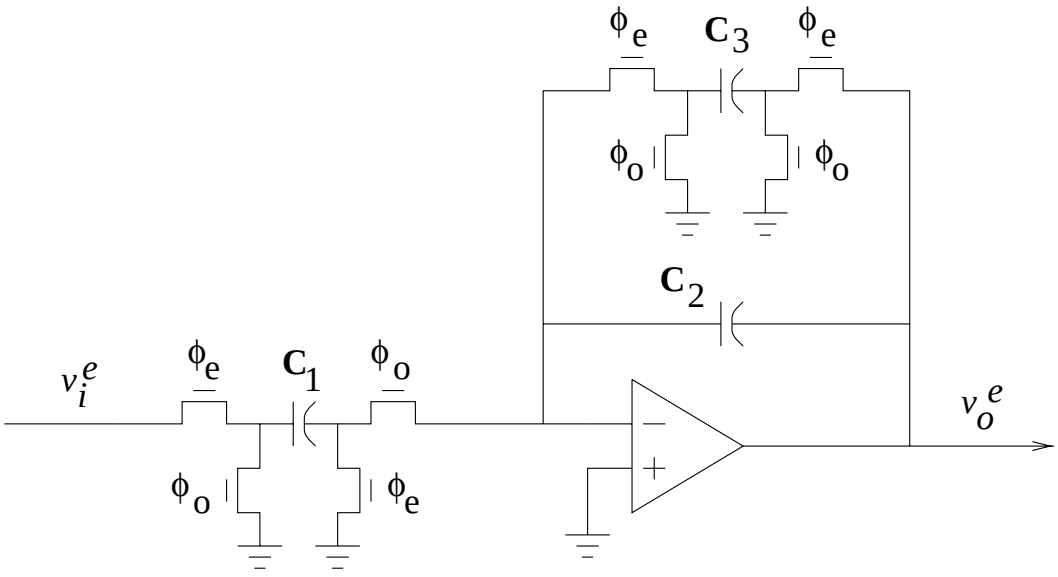


Figure 7.13: Switched capacitor implementation of damped integrator

phasing of the top switches. Similarly the configuration of Fig 7.12(b) will be denoted by the acronym EO to reflect the phasing of its top two switches. Using parasitic insensitive switches, Fig 7.13 shows a damped switched capacitor integrator which implements the following transfer function:

$$\frac{V_o^e}{V_i^e} = \frac{\frac{C_1}{C_2+C_3}z^{-1}}{1-\frac{C_2}{C_2+C_3}z^{-1}} \tag{7.15}$$

The subscript i in V_i^e refers to the input, the superscript e refers to the input being sampled on the even phase. The subscript o in V_o^e refers to the output, the superscript e refers to the output being sampled on the even phase. Note that the integrating capacitor C_2 is being updated on both odd and even phases and so the circuit is time varying. However as long as all inputs and outputs are sampled on the even phase the transfer function obtained on the odd phase is irrelevant. The circuit of Fig 7.13 will be denoted in short hand by the switch configuration EOEE/E with the last E after the slash representing that the output is sampled on the even phase. Fig 7.14 shows a full circuit of the complex integrator using the damped integrator of Fig 7.13. The additional subscripts r and j in this figure refer to the real and imaginary channels. Note that all inputs and output are sampled on the even phase. The switch configurations for the various paths are EOEE/E.

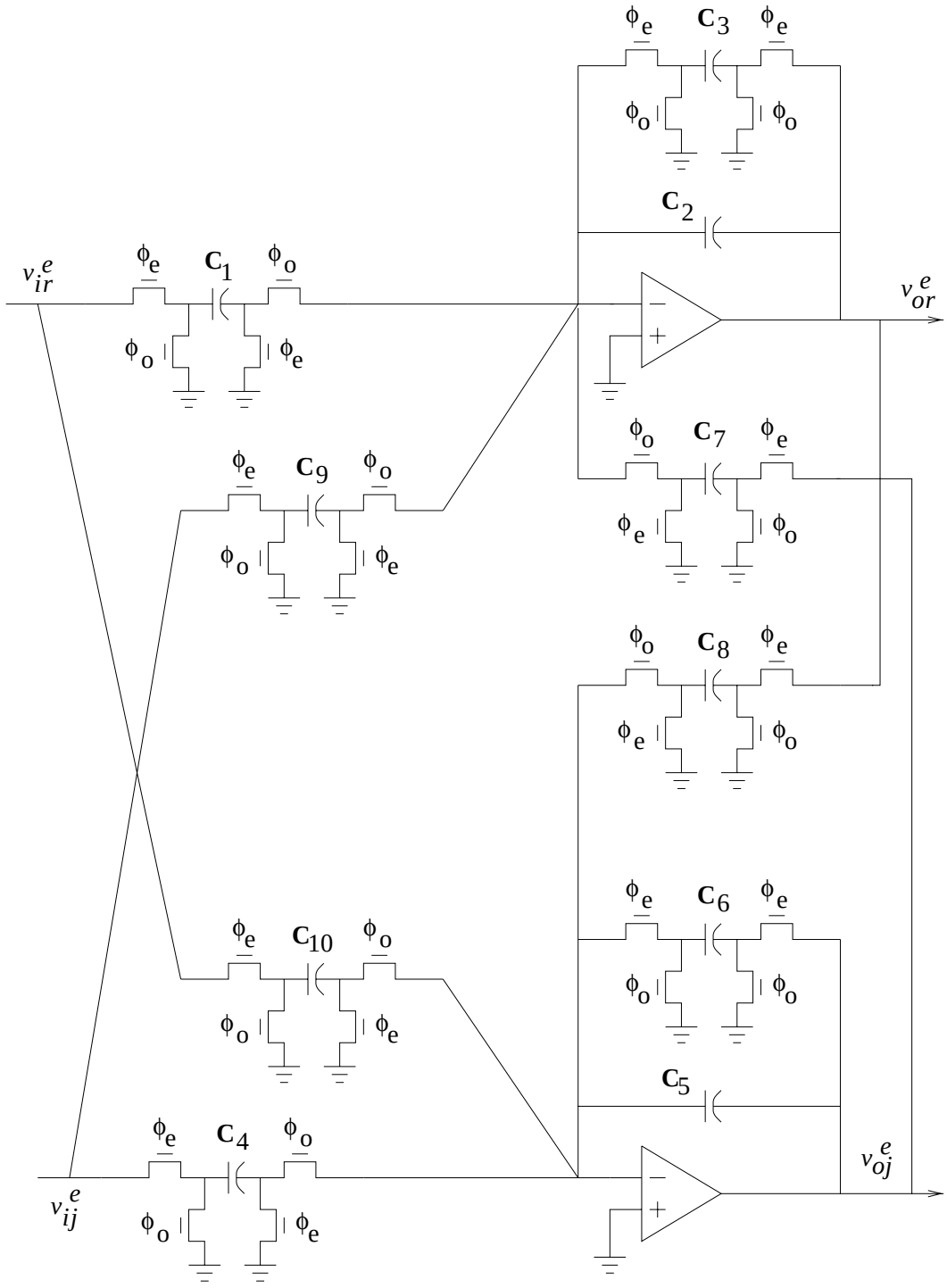


Figure 7.14: Switched capacitor implementation of complex integrator

7.3.2 Time Sharing for Complex Integrator

Now let us consider what is needed for a time shared realization for Fig 7.14. For the time shared implementation we have the following constraints [72]:

1. The different transfer function outputs, (in our case the real and imaginary transfer function outputs) should be sampled on different phases of the clock. Let us use E for the real channel output sampling and O for the imaginary channel output sampling.

2. The integrating capacitor for each transfer function should be updated on different phases. This phase should be the same as the phase chosen for output sampling for the transfer function. Consequently, the integrating capacitor for the real channel should be updated on the E phase and for the imaginary channel on the O phase.

For the moment if we assume no cross coupling between channels we can meet the above constraints by choosing the following switch configurations for the direct paths of the real and imaginary channels:

- real channel: E O E E E / E
  Input sampling: E phase
  Integrating capacitor update: E phase
  Output sampling: E phase
  The transfer function is of the form $\frac{gz^{-1}}{1-cz^{-1}}$ as desired.

- imaginary channel: O E O O O / O
  Input sampling: O phase
  Integrating capacitor update: O phase
  Output sampling: O phase
  The transfer function is of the form $\frac{gz^{-1}}{1-cz^{-1}}$ as desired.

The above configurations make use of the parasitic compensated toggle (PCT) switched capacitor circuit [74] shown in Fig 7.15 for use with the direct (non-cross coupled) path of the real channel. The transfer function is the same as in (7.15) It can be verified through an exhaustive search of all possible switch configurations that the time sharing constraints

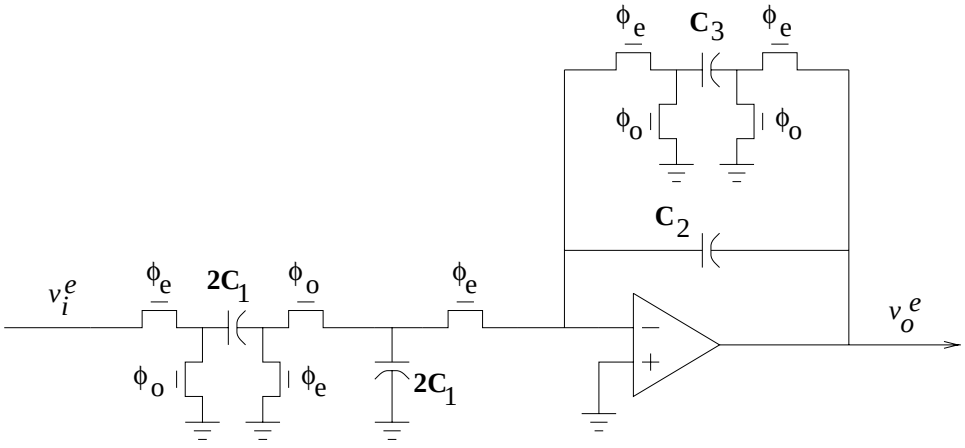


Figure 7.15: Parasitic compensated toggle switched capacitor damped integrator

necessitate the use of PCTs rather than the parasitic insensitive switches of the type shown in Fig 7.12.

Now let us consider the presence of cross coupling as required for our complex integrator. The imaginary channel input will be coupled to the real channel and so will the imaginary channel output. Both of these are sampled on the O phase so they could be coupled in the following way: O O E E E This way the integrating capacitor in the real channel is still updated on the E phase with respect to the imaginary channel cross couplings. The output of the real channel is, of course, still sampled on the E phase. With this we observe that the transfer function between the real channe output and say the imaginary channel input with the O O E E E configuration is of the form,

$$\frac{V_{or}^e}{V_{ij}^e} = \frac{gz^{-\frac{1}{2}}}{1 - cz^{-1}} \tag{7.16}$$

which is different from the desired transfer function by one half clock cycle. In fact, the transfer function from an E phased input to an output sampled on the O phase is always a multiple of a half clock cycle and so the need for realizing a z^{-1} delay for the cross coupled trasnfer functions conflicts with the need to have the real and imaginary trasnfer function outputs be sampled on different phases for time sharing. This is true even for the case $\omega_c = \frac{\pi}{2}$ where $c = 0$ and $s = 1$ in Fig 7.2. Consequently, based on the preliminary investigation conducted here it does not appear that time sharing is directly applicable to the complex integrator.

Chapter 8

Parallel Time Sliced Delta-Sigma A/D Conversion

8.1 Introduction

Slicing of the signal in time in the context of PCM conversion was discussed in Chapter 5. Time slicing is now discussed in the context of $\Delta\Sigma$ conversion. This $\Delta\Sigma$ architecture has been proposed in [75] to convert low-pass baseband signals. This chapter reviews the operation and performance of this architecture. In addition, we extend the idea of time slicing to include real band-pass as well as complex band-pass modulators. Simulations have been performed to validate this extension.

8.2 Time Sliced Low-pass $\Delta\Sigma$ Modulation

Fig 8.1 shows a block diagram of a block filter based time sliced delta-sigma A/D converter which has been proposed in [75]. The above architecture employs ideas of block filter theory to use P identical, *mutually cross coupled*, modulators with NTF $H_e(z)$ running in parallel at a sampling rate f_s to generate an equivalent overall modulator NTF $H_e(z)$ which runs at an equivalent sampling rate of $P \times f_s$. A more detailed block diagram is shown Fig 8.2 for a two channel system using 1st order modulators running in each individual channel. The existence of mutual cross-coupling is in contrast with frequency sliced conversion where

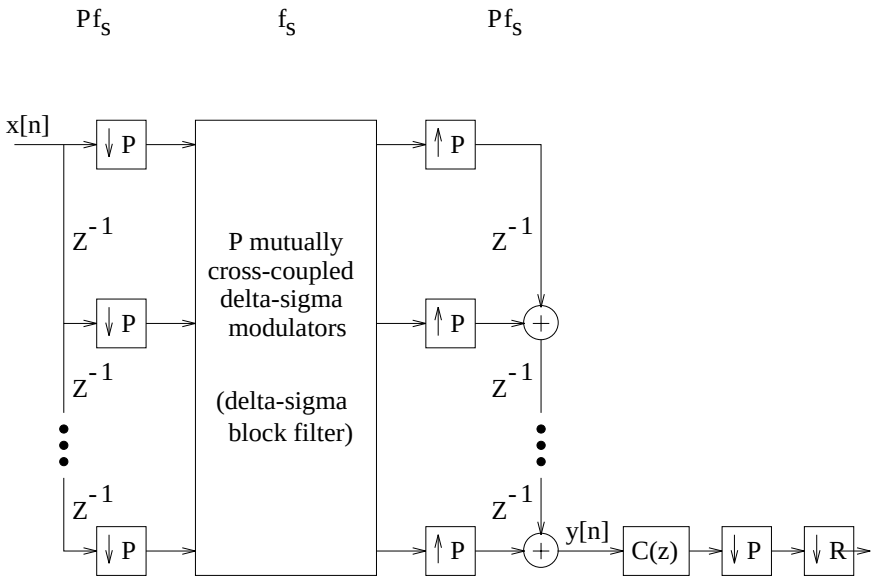


Figure 8.1: Block diagram of time interleaved delta-sigma modulator ADC

all the quantizers worked independently – the cross coupling terms are zero in the block filter theory context. If H_x and H_e are the STF and NTF of the individual modulators, the overall output is given by,

$$Y(z) = X(z)H_x(z) + E(z)H_e(z) \tag{8.1}$$

where,

$$E(z) = \sum_{p=0}^{P-1} E_p(z^P) z^{-(P-1-p)} \tag{8.2}$$

The overall error process is exactly the same as for the time interleaved PCM case. Note that downsampling the signal can result in aliasing but the choice of the cross coupling terms ensures that the aliasing is cancelled in the final output for the ideal system. In Chapter 3, we observed that for order L low-pass modulators, the SNR improves at a rate of $(6L + 3)$ dB for every doubling of the sampling rate. Therefore, if these modulators are used in a P channel time interleaved system, the SNR will improve at the rate of $(6L + 3)$ dB for every doubling of the number of channels, P, since using P channels is equivalent to increasing the sampling rate by a factor of P.

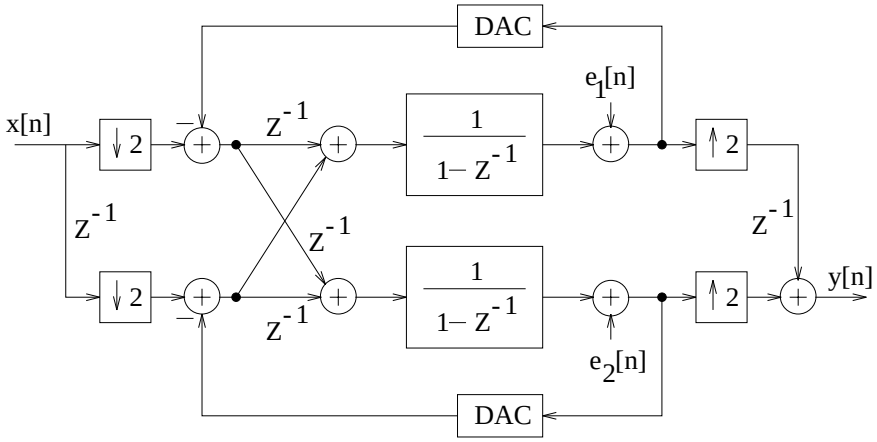


Figure 8.2: Two channel time interleaved conversion using 1st order ΔΣ modulators

8.3 Time Sliced Band-pass ΔΣ Modulation

We now extend the use of time interleaving to band-pass ΔΣ modulators. As we have seen in Chapter 6, the SNR of order L complex band-pass, or order $2L$ real band-pass ΔΣ modulators, improves at a rate of $(6L + 3)$ dB per octave increment in the sampling rate. Therefore, if these modulators are used in a P channel time interleaved system, the SNR will improve at the rate of $(6L + 3)$ dB for every doubling of the number of channels, P, since using P channels is equivalent to increasing the sampling rate by a factor of P.

In the band-pass architecture, there will again be mutual cross coupling among all the modulators. The structure of the modulator can be derived in the same manner as was done in [75] for the low-pass modulator. As before, the cross coupling will be based on the polyphase components of the modulator integrators. A Z domain transfer function, $H(z)$ can be written in terms of its P polyphase components, $H_p(z)$, $p = 0, 1, \ldots P-1$ as,

$$H(z) = \sum_{p=0}^{P-1} H_p(z^P) z^{-p} \tag{8.3}$$

When $H(z)$ is implemented in a time interleaved fashion, the computation of the time interleaved outputs in the P channels can be represented in matrix form as,

$$\overline{S}(z) = \overline{\overline{H}}(z)\overline{T}(z) \tag{8.4}$$

This matrix equation represents the block-filter box of Fig 8.1. $T(z)$ is a vector consisting of P internal inputs of the time interleaved channels. These P inputs are obtained from the

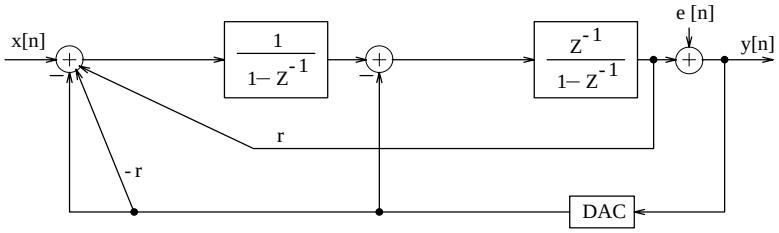


Figure 8.3: Second order real band-pass modulator obtained by modifying second order modulator of Fig 3.6.

outputs of the downsamplers shown in the left hand part of Fig 8.1. $S(z)$ is a vector of P internal outputs of the time interleaved channels. These P outputs then become inputs to the upsamplers in the right hand side of Fig 8.1. The matrix $\overline{\overline{H}}(z)$ is constructed from the polyphase components of $H(z)$ in such a manner so as to allow the overall transfer function of the system to be $H(z)$. This is done by constructing $\overline{\overline{H}}(z)$ to be a pseudocirculant matrix [59]. For a $P = 2$ channel system, (8.4) is expanded as,

$$\begin{bmatrix} S_0(z) \\ S_1(z) \end{bmatrix} = \begin{bmatrix} H_0(z) & H_1(z) \\ z^{-1}H_1(z) & H_0(z) \end{bmatrix} \begin{bmatrix} T_0(z) \\ T_1(z) \end{bmatrix} \tag{8.5}$$

where $S_0(z), S_1(z)$ are the internal outputs and $T_0(z), T_1(z)$ are the internal inputs of the $p = 0$ and $p = 1$ channel respectively.

Let us take a couple of examples to illustrate this process. Fig 8.3 shows a second order real band-pass modulator structure which we have obtained by modifying the second order low-pass modulator of Fig 3.6. The band-pass modulator's STF is $H_x(z) = z^{-1}$ and its NTF is $H_e(z) = 1 - (2 + r)z^{-1} + z^{-2}$ where r is a coefficient which sets the center frequency of the NTF. The coefficient $r = 2\left(\cos(\omega_c) - 1\right)$ where ω_c is the center frequency in radians.

For the $P = 2$ channel case, the polyphase components of the integrator $\frac{1}{1-z^{-1}}$ are $H_0(z) = \frac{1}{1-z^{-1}}$ and $H_1(z) = \frac{1}{1-z^{-1}}$. The expanded input/output relation of the two channels of this integrator are

$$S_0(z) = T_0(z)H_0(z) + T_1(z)H_1(z) = \frac{1}{1 - z^{-1}}\left[T_0(z) + T_1(z)\right] \tag{8.6}$$

$$S_1(z) = T_0(z)H_1(z)z^{-1} + T_1(z)H_0(z) = \frac{1}{1 - z^{-1}}\left[T_0(z)z^{-1} + T_1(z)\right] \tag{8.7}$$

Fig 8.4 shows the modulator of Fig 8.3 implemented as a 2 channel time interleaved system. The cross coupling structure for the integrator $\frac{1}{1-z^{-1}}$ as reflected in equations (8.6) and

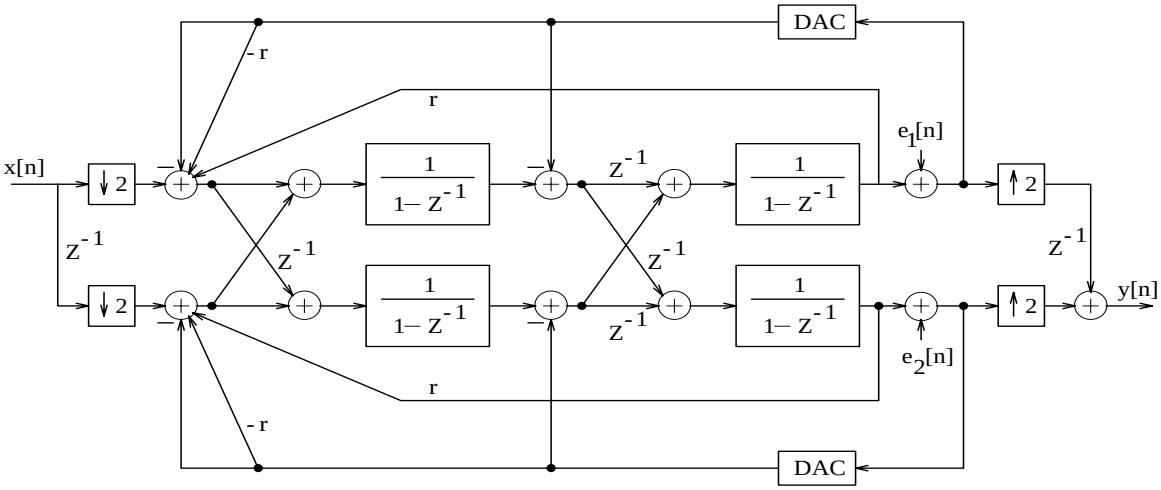


Figure 8.4: Time sliced second order real band-pass ΔΣ modulator obtained from modulator of Fig 8.3.

(8.7) is clearly visible in the left hand part of the figure. For the $P = 2$ channel case, the polyphase components of the second integrator $\frac{z^{-1}}{1-z^{-1}}$, of Fig 8.3, are $H_0(z) = \frac{z^{-1}}{1-z^{-1}}$ and $H_1(z) = \frac{1}{1-z^{-1}}$. The expanded input/output relation of the two channels of this integrator are

$$S_0(z) = T_0(z)H_0(z) + T_1(z)H_1(z) = \frac{1}{1-z^{-1}}\left[T_0(z)z^{-1} + T_1(z)\right] \tag{8.8}$$

$$S_1(z) = T_0(z)H_1(z)z^{-1} + T_1(z)H_0(z) = \frac{1}{1-z^{-1}}\left[T_0(z)z^{-1} + T_1(z)z^{-1}\right] \tag{8.9}$$

The cross coupling structure for the integrator $\frac{z^{-1}}{1-z^{-1}}$ as reflected in equations (8.8) and (8.9) is visible in the right hand part of the figure. Note that the number of r coefficients has doubled.

For complex band-pass modulators which employ complex integrators, there will be complex mutual cross couplings. Consider the complex integrator,

$$H(z) = \frac{z^{-1}e^{j\omega_c}}{1 - z^{-1}e^{j\omega_c}} \tag{8.10}$$

The P polyphase components of this integrator are

$$H_p(z) = \begin{cases} \frac{z^{-1}e^{jP\omega_c}}{1-z^{-1}e^{jP\omega_c}} & p = 0 \\ \frac{e^{jp\omega_c}}{1-z^{-1}e^{jP\omega_c}} & p = 1, 2 \ldots P-1 \end{cases}$$

For the $P = 2$ channel case, $H_0(z)$ and $H_1(z)$ are the only polyphase components. The

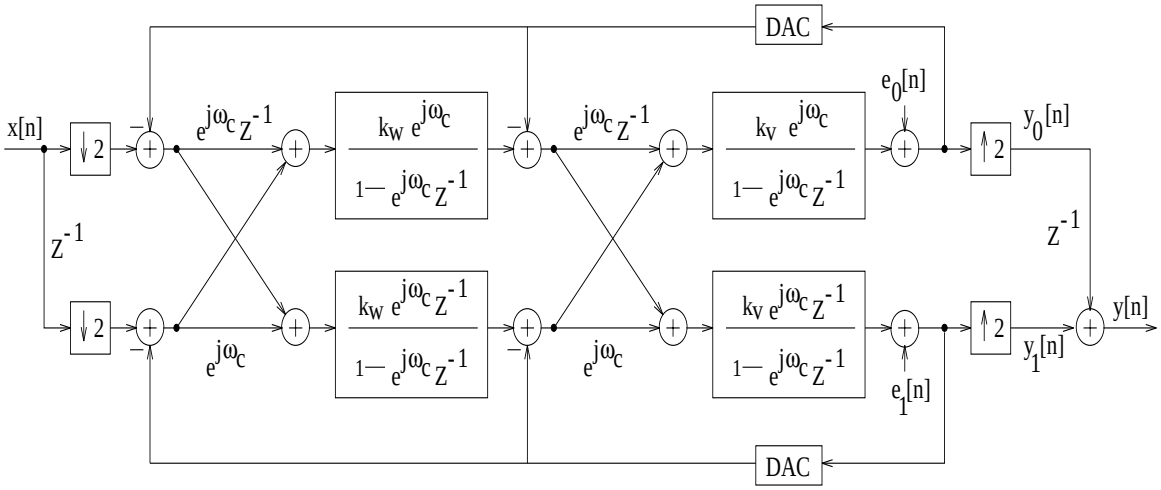


Figure 8.5: Time sliced second order complex BPΔΣ modulator. Signal lines are complex signals which have not been decomposed into real and imaginary parts.

expanded input/output relation of the two channels of the complex integrator are

$$S_0(z) = T_0(z)H_0(z) + T_1(z)H_1(z) = \frac{1}{1-e^{j2\omega_c}z^{-1}}\left[T_0(z)e^{j2\omega_c}z^{-1} + T_1(z)e^{j\omega_c}\right] \qquad (8.11)$$

$$S_1(z) = T_0(z)H_1(z)z^{-1} + T_1(z)H_0(z) = \frac{z^{-1}}{1-e^{j2\omega_c}z^{-1}}\left[T_0(z)e^{j\omega_c} + T_1(z)e^{j2\omega_c}\right] \qquad (8.12)$$

A block diagram of a two channel time sliced second order complex ΔΣ modulator is shown in Fig 8.5. Note that the signals here are all complex and have not been broken into their real and imaginary parts. When the real and imaginary parts are taken into account, this structure will become four physical channels.

8.4 Simulation Results

In order to validate the extension of time slicing to band-pass modulators, simulations were performed on the single channel second order real band-pass modulator of Fig 8.3 and the two channel time interleaved modulator of Fig 8.4. The quantizers were replaced with additive noise sources for the sake of making this validation. In Chapter 6, we observed that the SNR performance of an order $L = 2$ real band-pass NTF should increase at the rate of $(3L + 3) = 9$ dB per octave increment in the oversampling ratio. Therefore, the SNR of a time interleaved system using two modulators operating at the same sampling rate as the single channel system will be 9 dB better than that of the single channel system.

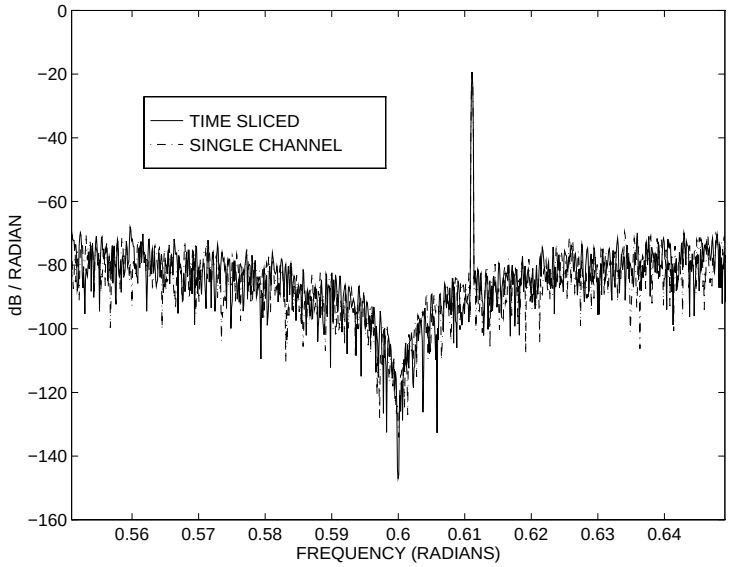


Figure 8.6: Power spectra of time sliced and single channel 2nd order band-pass delta-sigma modulators. $R = 128$, $\omega_c = 0.6$.

Equivalently, the SNR of a time interleaved system using two modulators operating at half of the sampling rate as the single channel system will be the same as that of the single channel system. This is verified by the simulations. Fig 8.6 shows the in-band power spectrum estimate of the output obtained by simulating the single channel system and the two channel time interleaved system such that the modulators of the two channel system operated at the half the rate as the single channel system. Note that the power spectral densities are virtually the same. The input used was a sinusoid near the center frequency of $\omega_c = 0.6$ radians. The oversampling ratio of the signal is $R = 128$ with respect to the sampling rate at the inputs of both systems. Table 8.1 documents the SNRs obtained for the two systems for several different oversampling ratios (R) and center frequencies ω_c. As expected, there is a good match between the two channel time sliced and single channel systems. In Chapter 9 we discuss the performance of time slicing delta-sigma systems in comparison with frequency sliced systems.

	R=32		R=128	
ω_c	**Time Slcd.**	**Sngl. Ch.**	**Time Slcd.**	**Sngl. Ch.**
0.2	42.32	41.87	58.99	60.31
0.6	32.77	32.96	49.92	50.57
1	29.68	29.31	47.14	47.09
$\frac{\pi}{2}$	28.03	28.05	45.85	44.99

Table 8.1: Simulation SNR (dB) results for time sliced (time slcd.) and single channel (sngl. ch.) 2nd order band-pass modulators. All quantizers have the same reference levels of 1 and -1.

Chapter 9

Parallel Frequency Sliced Delta-Sigma A/D Conversion

9.1 Introduction

Having looked at frequency and time sliced PCM conversion, and time sliced $\Delta\Sigma$ conversion, let us now consider frequency slicing for $\Delta\Sigma$ A/D conversion. This chapter first considers a sub-band coder structured system for $\Delta\Sigma$ modulators. It is shown that the architecture is not very useful. The thesis then proposes the use of a direct multi-band frequency slicing architecture for $\Delta\Sigma$ A/D conversion. This architecture, which we presented in [76], is examined and its SNR performance is analyzed. The thesis then proposes a reduced complexity filter bank, which we have described in [77], for use with the multi-band $\Delta\Sigma$ system. The thesis also provides an overall design procedure and examples for the high level design of multi-band $\Delta\Sigma$ systems. Finally, the practical performance of multi-band systems is discussed.

9.2 $\Delta\Sigma$ Sub-band Coder Structure

First, consider the effect of placing band-pass delta-sigma converters, with signal transfer function (STF) $H_{xp}(z)$ and noise transfer function (NTF) $H_{ep}(z)$, directly in a sub-band coder structure. A block diagram of this system is shown in Fig 9.1. The overall output is

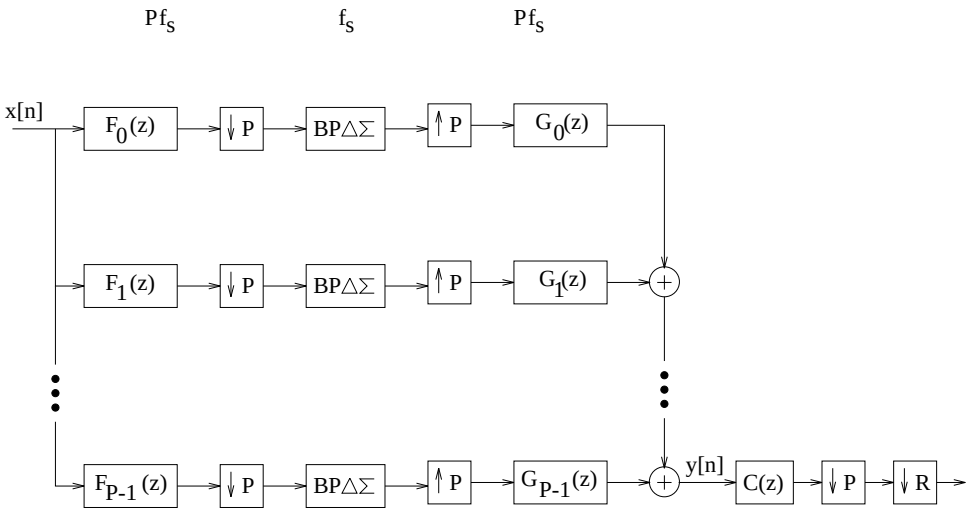


Figure 9.1: Block diagram of sub-band coder structure with ΔΣ quantizers

$Y(z)$,

$$Y(z) = \sum_{l=0}^{P-1} X(zW^l) \left(\frac{1}{P}\right) \sum_{p=0}^{P-1} H_{xp}(z) F_p(zW^l) G_p(z) + \sum_{p=0}^{P-1} E_p(z^P) H_{ep}(z^P) G_p(z) \qquad (9.1)$$

Looking at the noise component, we notice that the NTF $H_{ep}(z)$ is distorted and frequency scaled to $H_{ep}(z^P)$ relative to the signal at the same sampling rate. Since the attenuation characteristic of $H_{ep}(z^P)$ is much worse than that of $H_{ep}(z)$, this approach to frequency sliced ΔΣ modulation will not produce favorable results.

9.3 Direct Multi-band ΔΣ (MBΔΣ) Conversion

We now propose a direct multi-band ΔΣ which can be used to perform frequency sliced conversion employing ΔΣ converters. The architecture uses modulators which realize different band reject NTFs for different portions of the signal band. Each band (channel) is converted in parallel. We have designed a bank of FIR filters, which we described in [77], to attenuate the out-of-band noise for each band. In general, the bands need not be of equal size. The filter bank easily achieves perfect reconstruction of the signal component if the modulator STF is a simple delay. A general block diagram of the system architecture is shown in Fig 9.2 for P channels. $G_p(z)$ are the filters in the digital filter bank. The output

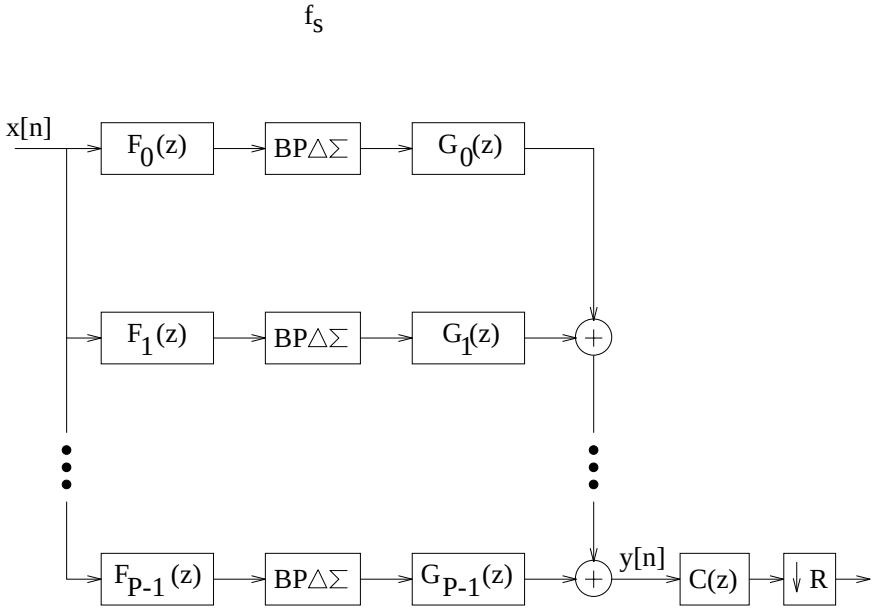


Figure 9.2: Block diagram of direct multi-band delta-sigma architecture

$Y(z)$ of the system prior to decimation is given by,

$$Y(z) = X(z) \sum_{p=0}^{P-1} H_{xp}(z) F_p(z) G_p(z) + \sum_{p=0}^{P-1} E_p(z) H_{ep}(z) G_p(z) \tag{9.2}$$

We will assume that $F_p(z) = 1$ and that the STFs are all simple delays with $H_{xp}(z) = z^{-K}$. Let us also assume that each channel consists of a band-pass modulator with an order L FIR complex noise transfer function with center frequency ω_{cp}. The output is then given by,

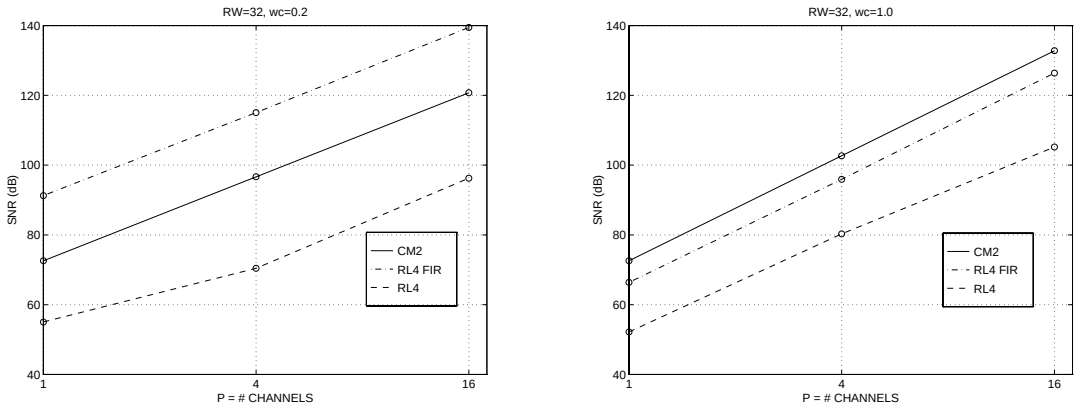
$$Y(z) = X(z) z^{-K} \sum_{p=0}^{P-1} G_p(z) + \sum_{p=0}^{P-1} E_p(z) G_p(z) \left(1 - z^{-1} e^{j\omega_{cp}}\right)^L \tag{9.3}$$

For the case of roughly equal sized bands, the signal band is $f_W = P f_B$ where f_B is the bandwidth quantized by each one of the P modulators. The in-band noise power at the output is P times the in-band noise power of any individual channel and is found by multiplying the right hand side of (6.4) by P to obtain,

$$\sigma_{ey}^2 = \frac{4P\sigma_e^2 \pi^{2L}}{2^{2L+1}(2L+1)} \left(\frac{2f_B}{f_s}\right)^{2L+1} \tag{9.4}$$

With $f_W = P f_B$, we obtain,

$$\sigma_{ey}^2 = \frac{4\sigma_e^2 \pi^{2L}}{2^{2L+1}(2L+1)} \left(\frac{2f_W}{f_s}\right)^{2L+1} \left(\frac{1}{P}\right)^{2L} \tag{9.5}$$



(a) (b)

Figure 9.3: Data from Tables 1-3 used to determine linearized SNR (dB) for a multi-band system using complex second order, real fourth order FIR (unstable), real fourth order IIR (stable) NTFs for $R = 32$. SNR is plotted vs number of channels, $R_B = PR_W$. Notice the 12 dB per octave slope. (a) $\omega_c = 0.2$ radians (b) $\omega_c = 1.0$ radians

From this we can infer that for every doubling of the number of channels, P, the SNR improves at a rate of $(6L)$ dB. The SNR of modulators employing order $2L$ real band-pass NTFs also improves at a rate of $(6L + 3)$ dB for every doubling of the oversampling ratio per band. Consequently the SNR of a multi-band system using these modulators will also improve at a rate of $(6L)$ dB for every doubling of the number of channels, P. As an example, let us consider using the complex second order and real fourth order NTFs of Chapter 6 in a multi-band system. We can look at the SNR performance of the system as a function of the number of channels or bands, P. Doing this for the case where the oversampling ratio of the overall signal band is $R = \frac{R_B}{P} = 32$, we obtain the plots shown in Fig 9.3. We observe, as expected, that the SNR improves at a rate of 12 dB per octave increment in P for both $cm2$ and $rl4$ NTFs but the absolute performance of the $cm2$ NTF is better than that of the $rl4$ NTF because the pole contributions of the $rl4$ NTF degrades in-band attenuation as was discussed in Chapter 6.

We have observed that the direct multi-band frequency sliced approach provides a SNR improvement with respect to the number of channels P at the rate of $(6L)$ dB per octave increment in the number of channels for the types of NTFs mentioned. The time sliced method, however, displayed $(6L+3)$ dB improvement in performance per octave increment in the number of channels. The reason for this 3 dB difference is, of course, the same as that

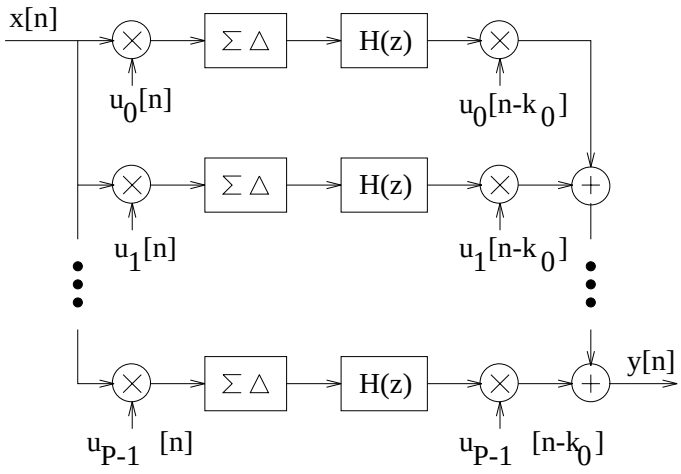


Figure 9.4: Hadamard modulated delta-sigma A/D conversion system

for PCM A/D conversion – the frequency sliced approach is unable to reduce the amplitude distribution of the signal seen at the input of each ΔΣ modulator. The averaging system, which in the PCM case performed as well as the time sliced system, is obviously not directly applicable to ΔΣ modulators because of the filtering and noise-shaping involved.

9.4 Other Transform Sliced Approaches

Before discussing the design of the direct multi-band frequency sliced architecture in further detail, let us briefly describe a frequency like parallel ΔΣ system which has appeared in the literature [78]. Slicing in this architecture is performed in the Hadamard domain. Here, each channel contains a Hadamard modulator, $u_k[n]$, which is a time domain ± 1 sequence that is multiplied by the input in the time domain. This operation is called Hadamard modulation. The Hadamard sequence is obtained by repeating the rows of the Hadamard transform matrix. The Hadamard modulated sequence of each channel is then converted using a standard ΣΔ modulator. The output of each ΔΣ modulator is filtered to attenuate out-of-band high frequency noise and again multiplied by a Hadamard sequence before all the channel outputs are added to provide the final output. The block diagram of the system is shown in Fig 9.4 where k_0 is assumed to be the delay in the STF of the ΔΣ modulators. It has been shown that the overall effect of the system on the signal is to filter it with a

subset of the filter coefficients of $H(z)$. The relevant $H(z)$ coefficients affecting the signal can be chosen such that the STF experiences a delay while the other coefficients can be chosen to maximize the quantization noise attenuation. The quantization error does not see the first Hadamard sequence and the effect of the second Hadamard sequence modulation is to frequency shift the quantization error power spectral density. Using standard Lth order low-pass $\Delta\Sigma$ modulators, the SNR using this approach improves by $(6L)$ dB per octave increment in the number of channels P.

There are two differences between this and the multi-band frequency approach we have proposed. (1) The multi-band frequency slicing is done in the Fourier domain whereas the Hadamard approach is performed in the Hadamard domain. (2) The frequency slicing in the multi-band case is achieved using different modulator NTFs frequency responses whereas the Hadamard domain slicing is achieved with time domain multiplication of the input signal with Hadamard sequences. An analogous scheme could be developed for the Fourier domain slicing using time domain multiplication by complex exponentials or only sines or cosines. The idea would be to demodulate different portions of the signal spectrum to base-band, perform $\Delta\Sigma$ conversion using a low-pass $\Delta\Sigma$ ADC and then remodulate the various parts of the spectrum to their original location. A digital filter bank would still be used to perform out-of-band quantization attenuation and reconstruction of the signal. The main disadvantage of Fourier multi-band frequency slicing based on time domain multiplication is the difficulty of implementing high quality analog multipliers compared with implementing NTF coefficients in terms of capacitor ratios in a switched capacitor technology. An exception to these would be the single band case for a band-pass $\Delta\Sigma$ with $\omega_c = \frac{\pi}{2}$. This indeed has been done in [79].

9.5 Multi-band $\Delta\Sigma$ Digital Filter Bank Design

This section provides a discussion of the design of the digital filter bank needed by a frequency sliced direct multi-band $\Delta\Sigma$ system to attenuate out-of-band quantization noise and to reconstruct the signal. We propose a design procedure which allows the filter bank to be designed with reduced computational complexity despite the fact that the bandwidths

of the filter are very narrow for large oversampling ratios. Next we extend the use of this procedure to ΔΣ modulators with complex NTFs. Finally, in Appendix C we show that if a sub-band coder follows the multi-band ΔΣ system, the digital filter complexity of the overall system can be reduced by integrating the multi-band ΔΣ filter bank with the filters of the sub-band coder.

9.5.1 Differences of Linear Phase Low-pass Prototypes

Up to now we have considered the signal bandwidth f_W to be broken up into P equal parts of width f_B. The design procedure accomodates **non-uniform** bands considering that f_W might be broken into P channels in a non-uniform fashion, let us define for convenience f_C to be the smallest of these bands. This provides a frame of reference in choosing transition band sizes for the filter design procedure to be described. For a given sampling rate, observe that having non-uniform bands means that not all bands have the same SNR or resolution since the SNR is directly related to the oversampling ratio which is different for the various bands. This variable resolution could be taken advantage of in certain applications where some bands are considered to be more important than others.

The filter bank transfer functions $G_p(z)$ can be designed from a set of P prototype low-pass linear phase FIR filters $H_p(z)$, all of the same order. The design can be done so that the overall response of the filter bank is a linear phase low-pass response. The filter H_p has a nominal cutoff frequency f_{p+1}. For a low-pass signal, the first band has a low-pass response, and we have $G_0(z) = H_0(z)$. For the remanining bands for which we desire a band-pass response, the filter G_p for the band with nominal band edges $[f_p, f_{p+1}]$ is obtained as the difference between the two low-pass prototype responses H_{p-1} and H_p with nominal cutoff frequencies f_p and f_{p+1} respectively.

$$G_p(z) = H_p(z) - H_{p-1}(z), p \neq 0 \tag{9.6}$$

The overall response, given by (9.7), is low-pass with nominal cutoff frequency f_W which is the total signal bandwidth.

$$\sum_{p=0}^{P-1} G_p(z) = H_{P-1}(z) \tag{9.7}$$

For $k \neq P-1$, the H_p have been designed with passband and stopband frequencies $[f_{p+1} - t * f_C, f_{p+1} + t * f_C]$ where $0 < t < 1$ is a width factor of the transition-band. The introduction of the term f_C now makes sense as the transition band widths of the filters are related to the smallest band. Since the overall filter bank response is the same as that of H_{P-1}, its passband and stopband frequencies have been chosen as $[f_W, f_W + 2 * t * f_C]$ to prevent droop in the signal band f_W. By construction, the signal is reconstructed perfectly and low-pass filtered with the response of H_{P-1}.

9.5.2 Computation Reduction With Frequency Scaling

A direct design of the H_p using the Parks-McClellan algorithm will lead to filter orders in excess of one thousand according to the formulae in [52] since the H_p have extremely narrow transition bands and are also narrowband relative to f_s. To make the filter order manageable, we use the simplest frequency scaling technique proposed in [80] to design the H_p. Let f_E be some excess bandwidth factor. The need for and meaning of this will be clear shortly. Typically, f_E will be a small portion of the signal bandwidth f_W. With a normalized or scaled half sample rate bandwidth of 0.5, the normalized bandwidth of the signal plus the excess bandwidth is $f_{SEn} = \frac{0.5}{f_W + f_E}$. Normalized or scaled versions of H_p, which will be called T_p, are now designed on this normalized frequency scale. Each T_p has nominal cutoff frequency $f_{pn}, p = 0 \ldots P$. The T_p, $p \neq P-1$, are designed with passband and stopband frequencies $[f_{pn} - t * f_{Cn}, f_{pn} + t * f_{Cn}]$ while T_{P-1} is designed with passband and stopband frequencies $[f_{Wn}, f_{Wn} + 2 * t * f_{Cn}]$. The T_p can be designed using the Parks-McClellan filter design program [81]. The normalized prototype low-pass filters are then frequency scaled to obtain the desired low-pass prototype filters:

$$H_p(z) = T_p(z^U) \tag{9.8}$$

where U =round$(\frac{0.5 f_s}{f_W + f_E})$. Note that if there is no excess bandwidth, U is simply the oversampling ratio $R_W = \frac{f_s}{2 f_W}$. The normalized nominal band edges frequencies f_{pn} now scale to the correct locations f_p. The H_p's are now used in (9.6) to generate the G_p's. It is important to note that although the order of the H_p's and hence the G_p's is $L_1 U$, only $L_1 + 1$ of the total $L_1 U + 1$ coefficients are non zero.

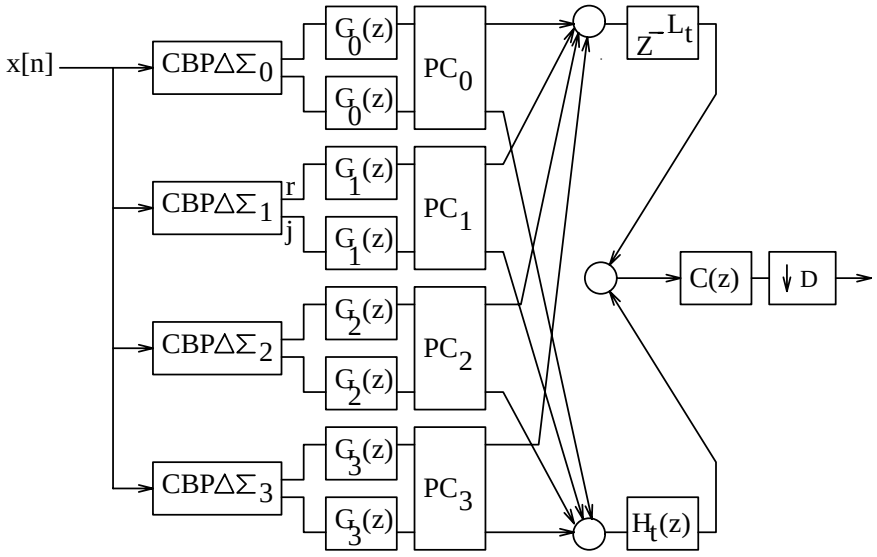


Figure 9.5: Multi-band delta-sigma architecture for complex modulators

A couple of other observations are in order. Note that even though the H_p's are equiripple, the G_p's are not guaranteed to be equiripple. Also, frequency scaling of the spectrum will cause images of the filter responses H_p to appear. The band edge of the first image will be at frequency $f_W + 2f_E$. In general, if the upper cutoff frequency of H_{P-1} is chosen as $f_W + 2 * t * f_C$ then to prevent any interference between it and its first image we require $f_W + 2 * t * f_C < f_W + 2f_E - 2t * f_C$ i.e. we must choose $f_E > 2tf_C$.

9.5.3 Extension to Use With Complex ΔΣ Modulators

We now directly apply the above filter design procedure to the case where the modulators in the multi-band system are complex. This is done by designing the filter bank assuming the modulators are real and simply filtering the real and imaginary components of each channel separately with the filter for that channel. This is, of course, possible because filtering a complex signal is equivalent to filtering the real and imaginary components with the filter and adding up the resulting outputs. This architecture is shown in Fig 9.5 for $P = 4$ complex channels or eight physical channels. The blocks labeled **PC** are the phase compensation blocks described in Chapter 6 and which is shown in detail in Fig 6.7. The filters to attenuate out-of-band quantization noise operate on the single bit output of the

modulators and so no hardware multipliers are required. They can also be designed with reduced computational complexity as discussed above.

9.5.4 Filter Bank Design Involving IIR STFs

Although signal reconstruction is not difficult in the case of the complex second order NTF whose signal transfer function (STF) is a simple delay, it is not trivial for general band-pass IIR STFs which are produced with real fourth order NTFs (and which would also be produced with higher order complex NTFs). This is true even if the STFs have in-band linear phase. However, optimizations of the sort performed in [66] using filtorX [67] could perhaps be generalized to jointly optimize several real STFs in a multi-band system.

9.6 Multi-band $\Delta\Sigma$ Design Procedure

In this section a design procedure is provided for the high level design of a multi-band $\Delta\Sigma$ A/D system. Design examples are provided in the next section to illustrate the procedure.

- f_W = signal bandwidth
- SNR = desired signal to noise ratio
- $f_{s,mx}$ = maximum sampling frequency allowed by technology

For various transfer functions and assuming uniform bandwidth channels (of course, the design procedure is able to accomodate non-uniform bandwidth channels if desired)

1. Plot curves of f_s vs number of physical channels (P for real NTFs, $2P$ for complex NTFs) needed to obtain SNR dB of resolution.
   If the form of the NTF is known (e.g. $rl2$, $cm2$, $rl4_{fir}$):
   - Use linearized analytical expressions such as (9.5) or frequency domain FFT based numerical analysis on the linearized transfer functions.
   - For each P, compute f_s required directly from analytical expression or iterate for various f_s up to $f_{s,mx}$ until f_s is found which yields SNR dB of resolution to within some tolerance.

else (the exact form of the NTF is not known e.g. $rl4$):
In this case additional work is required. For each P:

- Compute oversampling ratio per band R_B
- Compute P different center frequencies for the P bands $\omega_c = 2\pi\frac{f_c}{f_s}$ for the P NTFs
- Optimize the P NTFs at each ω_c using software such as filtorX to optimize the pole placement with respect to SNR and stability.
- Compute f_s required to obtain SNR dB of resolution to within some tolerance.

2. Based on f_s vs number of physical channel tradeoff, choose transfer function with best performance.
3. For best transfer function choose an "optimum" (f_s, P) pair.
4. Perform filter bank design.
   - For various t, plot SNR vs filter order L_1.
     This needs to be done using frequency domain based linearized analysis which weights the noise transfer function performance with the frequency response of the filter bank.
   - Choose t yielding best SNR vs L_1 curve.
   - Choose "optimum" L_1 based on point of diminishing returns on SNR vs L_1 curve.
5. Perform $C(z)$ decimator design.
6. Compile arithmetic unit count for filter bank and decimator.

9.7 Multi-band $\Delta\Sigma$ Design Examples

The design procedure is illustrated below with design examples which consider the following NTFs:

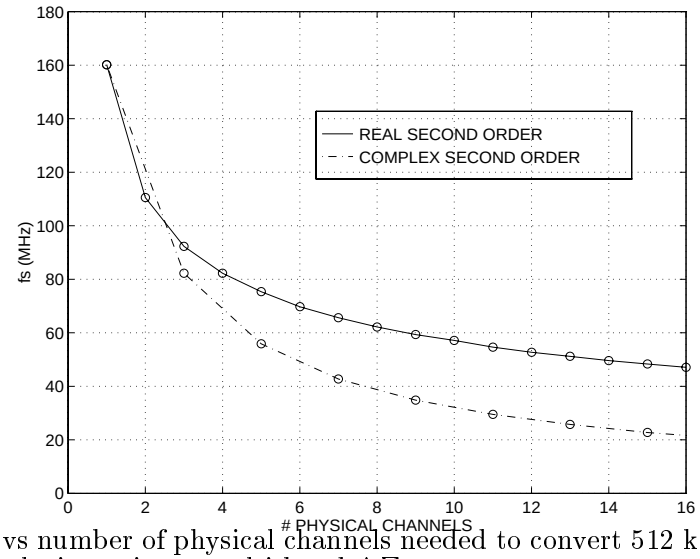


Figure 9.6: f_s vs number of physical channels needed to convert 512 kHz signal bandwidth to 16 bits of resolution using a multi-band $\Delta\Sigma$.

1. A multi-band system employing real second band-pass ($rl2$) NTFs. The band-pass modulator's STF is $H_x(z) = z^{-1}$ and its NTF is $H_e(z) = 1 - (2 + r)z^{-1} + z^{-2}$ where r is a coefficient which sets the center frequency of the NTF. The coefficient $r = 2\left(\cos(\omega_c) - 1\right)$ where ω_c is the center frequency in radians. The modulator structure for this NTF was shown in Fig 8.3.

2. A multi-band system employing the complex second order band-pass ($cm2$) NTFs which have been discussed throughout Chapters 6 and 7. The STF after phase compensation is $H_x(z) = z^{-2}$ and the NTF is $H_e(z) = \left(1 - z^{-1}e^{j\omega_c}\right)^2$.

The design specifications for this example are to convert a signal bandwidth of 512 kHz to 16 bits of resolution i.e. to 98 dB SNR for a full scale sinusoidal input. The $f_{s,max}$ will be considered to be 60 MHz so the f_s values chosen should be less than this. As per step 1 of the procedure we first plot the f_s vs P tradeoff which plots the sampling frequency needed vs the number of physical channels P needed to obtain 16 bits of resolution for the 512 kHz bandwidth. This is shown in Fig 9.6. We observe that for the $rl2$ modulator $P = 16$ channels are needed to obtain a f_s of about 48 MHz. For the $cm2$ modulators, it was found that $P = 4$ i.e. 8 physical channels were needed so the SNR for the $f_W = 512$ kHz signal band would be more than 98 dB with a nominal sampling frequency of about $f_s = 48$ MHz. Clearly the $cm2$ modulator performs better. However, we will illustrate further aspects of

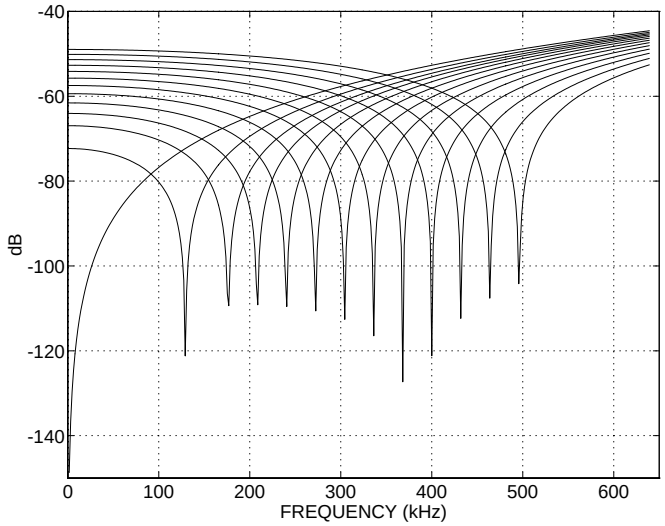


Figure 9.7: Noise transfer functions $|H_{ep}(f)|$ for design example

the design procedure with both of these modulators.

9.7.1 Design Example #1: $rl2$ NTF

We have observed that this NTF requires $P = 16$ channels to meet the specification with nominal $f_s = 48$ MHz. Since $f_W = 512$ kHz, each channel had a nominal bandwidth of 32 kHz. Although the initial design calculation suggested that $P = 16$ channels were required, it was found that reducing the number of channels to $P = 13$ where the first channel is of width 96 kHz, the next of width 64 kHz and the rest of width 32 kHz caused a negligible loss in SNR. This is because the real NTFs at lower center frequencies performed better than the NTFs at higher frequencies as is clear from the NTF plots shown in Fig 9.7 and from our earlier observations in Chapter 6. The actual design was done with a slightly higher than the nominal 48 MHz; $f_s = 52.224$ MHz was used.

Now let us describe the digital filter bank design for this example. Here, f_E was chosen to be $2f_C = 64$ kHz. Consequently, U turned out to be 45. The linearized model SNR, weighted by the filter transfer functions corresponding to the G_p, was used to determine the dependence of SNR on filter order L_1 of the T_p and transition-band factors t.

If the transition bands are made narrower by decreasing t, the approximate minimum

t	sba	L_1	L_{1e}	L_{1d}	$L_1 \times U$
0.90	46	50	47	1745	2250
0.75	46	60	56	2093	2700
0.50	38	70	63	2656	3150
0.25	44	160	154	6038	7200

Table 9.1: Minimum L_1 needed for 16 Bit Performance vs t ($U = 45$)

filter order, L_1, required to achieve the same resolution and bandwidth actually increases as shown in Table 9.1. This is because as long as the transition band is not extremely large (at most $1.8f_C$ in our examples), it is more important to have a good final stop band attenuation (sba) rather than a narrow transition band because the transition band affects the noise over only a relatively small region whereas the stop band attenuation affects the noise over a larger region. The sba values in Table 9.1. are for the prototypes H_p. The column L_{1e} shows the filter order estimates for $T_p(z)$ based on equation (3.63) in [52] while the column L_{1d} shows the filter order estimates based on (3.65) in [52] for a direct design of the $H_p(z)$ filters. A plot of L_1 versus SNR is given in Fig 9.8 for $t = 0.9$. It is seen that a minimum order of 40 is required to convert a signal power of 0.5 to 98 dB (16 bits) resolution with quantizer error powers of 0.38. To be conservative we will choose L_1 corresponding to a $\simeq$ 99.5 dB linearized model SNR.

For the above optimum parameters of $t = 0.9$ and $L_1 = 50$, the set of H_p transfer functions which were obtained is shown in Fig 9.9. The excess bandwidth factor is apparent in Fig 9.9 and note that the cutoff frequency of the image of H_{P-1} is at 640 kHz ($512+2\times64$ kHz). The G_p transfer functions corresponding to the H_p in Fig 9.9 are shown in Fig 9.10. Again note that the first image of G_{P-1} appears starting at 640 kHz.

Since the output of the modulators is single bit, no full hardware multipliers are required in the filter bank. The computation required by the filter bank is only L_1 additions per channel at the initial sampling rate f_s. The digital hardware required per channel is L_1 adders and L_1U delay registers which must operate at a rate of f_s.

The filter $C(z)$ is used to attenuate the images as well as any quantization noise beyond

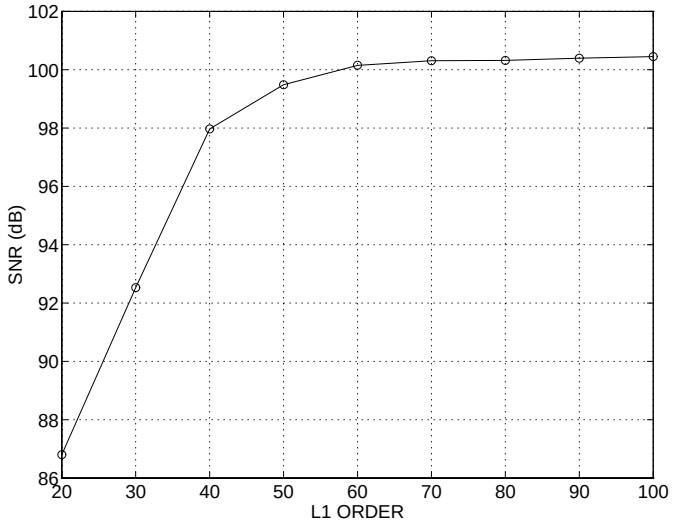


Figure 9.8: SNR vs filter order L_1 for $t = 0.9$

the signal band. $C(z)$ also needs to have a narrow transition band. Let us choose this to be f_B from f_W to $f_W + f_B$ and require enough attenuation by the frequency $f_W + f_B$ so as not to degrade the desired in-band signal to noise ratio. Because it precedes a downsampler, the filter complexity of $C(z)$ can be reduced significantly by implementing it as a multi-stage decimator with the filter for each stage of decimation being implemented efficiently with a polyphase structure [82]. The decimation filter can be designed without any transition band aliasing.

For the design example with $f_P = 512$ kHz, f_s=52.224 MHz, the downsampling factor is chosen as $D = 48$ instead of the full oversampling ratio of $R = \frac{f_s}{2f_W} = 51$. This brings the final sampling rate to 1.088 MHz which is $2 \times (f_W + f_B)$ and just above the Nyquist rate which is $2 \times f_W$ kHz = 1.024 MHz. A 150 dB stopband attenuation for the decimation filters will preserve a 16 bit noise floor. A two stage 24:2 decimation with 150 dB minimum stopband attenuation for the filters requires filter orders of 40 and 400 for the first and second stage of decimation respectively. The first stage filter can be augmented with 7 zero coefficients to result in a 48 tap filter which can be implemented as a polyphase structure thereby allowing the hardware to operate at a rate of $\frac{f_s}{24}$ and also allowing the number of registers to be reduced from 40 to 48/24=2. The second stage can also be implemented as a polyphase structure which will operate at $\frac{f_s}{48}$ with 201 registers.

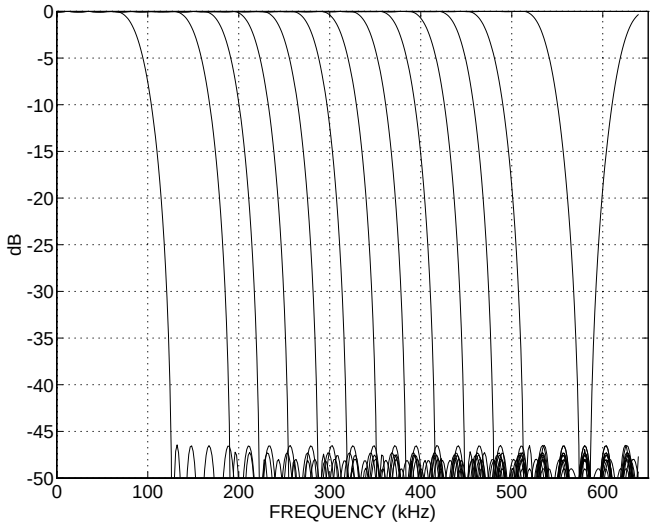


Figure 9.9: Low-pass prototype transfer functions $|H_p(f)|, t = 0.9, L_1 = 50$

9.7.2 Design Example #2: $cm2$ NTF

Actually, in this case, the modulator used in the first channel of the system was a standard second order real modulator and so only one physical channel was needed for the first of the $P = 4$ bands. So a total of 7 instead of 8 physical channels are needed to implement the $P = 4$ channel multi-band system.

Here the signal band was broken into uniform sizes of $f_C = 128$ kHz each. Fig 9.11 shows the standard real second order NTF for the first band and complex second order NTFs for the other bands.

The design procedure produced a value of $U = 34$ with an excess bandwidth of $2f_C = 256$ kHz. In performing the filter design we found that a mere filter order of $L_1 = 20$ was sufficient to meet the SNR specification for $t = 0.9, 0.75, 0.5$. For $t = 0.25$, $L_1 = 40$ or more is needed. The reason that lower filter orders can be used here than for the $rl2$ modulator is that there are fewer channels here – hence the oversampling ratio per band is lower for the $cm2$ system. The lower oversampling ratio means that the filters are not as narrow band and can achieve the same attenuation performance with a lower order. For this complex second order design example, Fig 9.12 shows the low-pass prototype transfer functions obtained with prototype filter order of $L_1 = 20$ and transition band factor of $t = 0.9$. Note the

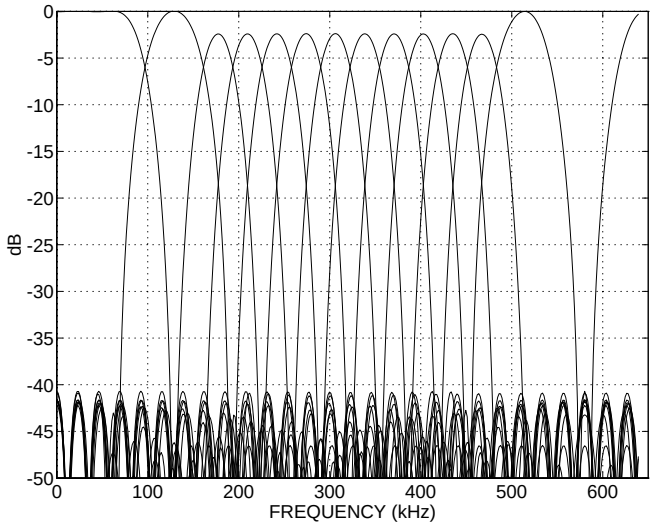


Figure 9.10: Filter bank transfer functions $|G_p(f)|, t = 0.9, L_1 = 50$

first image of H_3 appearing such that its cutoff frequency is around $f_W + 2f_E = 1024$ kHz. Finally, Fig 9.13 shows the filter bank transfer functions obtained from the low-pass prototypes. Again, note the first image of G_3 apearing at 1024 kHz.

Again, since the output of the modulators is single bit, no full hardware multipliers are required in the filter bank. The computation required by the filter bank is only L_1 additions per channel at the initial sampling rate f_s. The digital hardware required per channel is again L_1 adders and L_1U delay registers which must operate at a rate of f_s. For $L_1 = 20$ and $U = 34$, 20 additions per channel are required and 680 registers are required for each of the 7 physical channels.

9.8 Practical Performance of Multi-band ΔΣ

9.8.1 SNR Degradation of the Multi-band System

The SNR degradation of a single band-pass channel can be directly related to the SNR degradation of a P band multi-band frequency sliced system provided that the SNR degradation is such that (1) the added noise at the output of the modulator is relatively white and (2) all the modulators will contribute approximately the same added noise power at the output of the system. The added noise will be approximately white if the SNR degradation

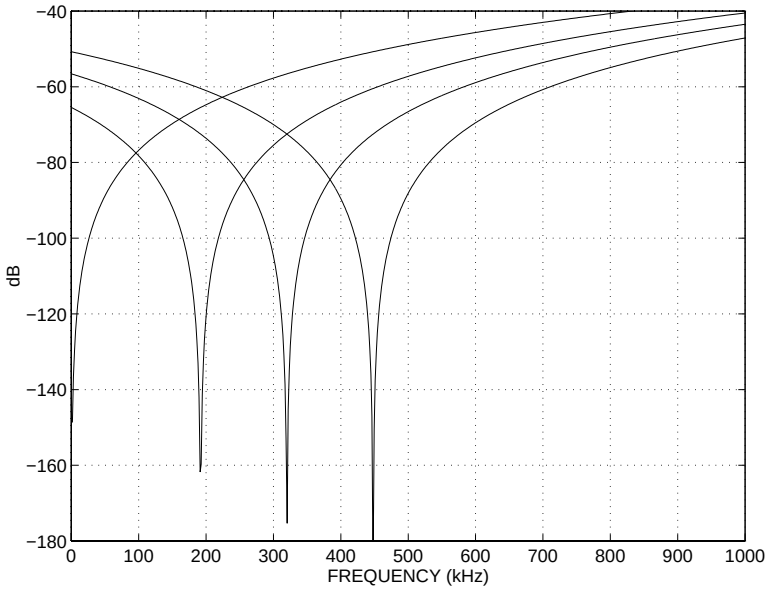


Figure 9.11: Noise transfer functions $|H_{ep}(f)|$ for complex ΔΣ design example

is not too large. For the $cm2$ modulator we know that the performance does not change with respect to center frequency and so, in a multi-band system, each band will have about the same added noise power due to implementation imperfections. For the $rl4$ NTF, as long as the oversampling ratio of the overall signal band is not too small and the number of bands is not too large, the transfer functions of the bands will be similar and the added noise at the output of each band will also be similar.

Given the above assumptions, the SNR loss in dB for the overall system is the same as that for any particular channel. This can be seen as follows.

Let σ_n^2 be the in-band quantization noise from the modulator of any one band. Let the signal power for that band be σ_u^2. The SNR for that band is $\frac{\sigma_u^2}{\sigma_n^2}$. Let the noise power with implementation imperfections be $\sigma_n^2 + \Delta\sigma_n^2$. The new SNR is $\frac{\sigma_u^2}{\sigma_n^2+\Delta\sigma_n^2}$ and the loss in SNR in dB, ΔSNR, is obtained from $\frac{\sigma_u^2}{\sigma_n^2} / \frac{\sigma_u^2}{\sigma_n^2+\Delta\sigma_n^2} = \frac{\sigma_n^2}{\sigma_n^2+\Delta\sigma_n^2}$.

Let the total signal power be σ_x^2. The total in-band noise of the overall system without implementation imperfection is $\sigma_{nt}^2 = P\sigma_n^2$. The corresponding system SNR is $\frac{\sigma_x^2}{P\sigma_n^2}$. With implementation imperfections, the total noise power becomes, $\sigma_{nt}^2 = P(\sigma_n^2 + \Delta\sigma_n^2)$. With implementation imperfections, the new system SNR is $\frac{\sigma_x^2}{P(\sigma_n^2+\Delta\sigma_n^2)}$ and the loss in SNR in dB, ΔSNR, is obtained from $\frac{\sigma_x^2}{P\sigma_n^2} / \frac{\sigma_x^2}{P(\sigma_n^2+\Delta\sigma_n^2)} = \frac{\sigma_n^2}{\sigma_n^2+\Delta\sigma_n^2}$. Thus, the measured SNR loss

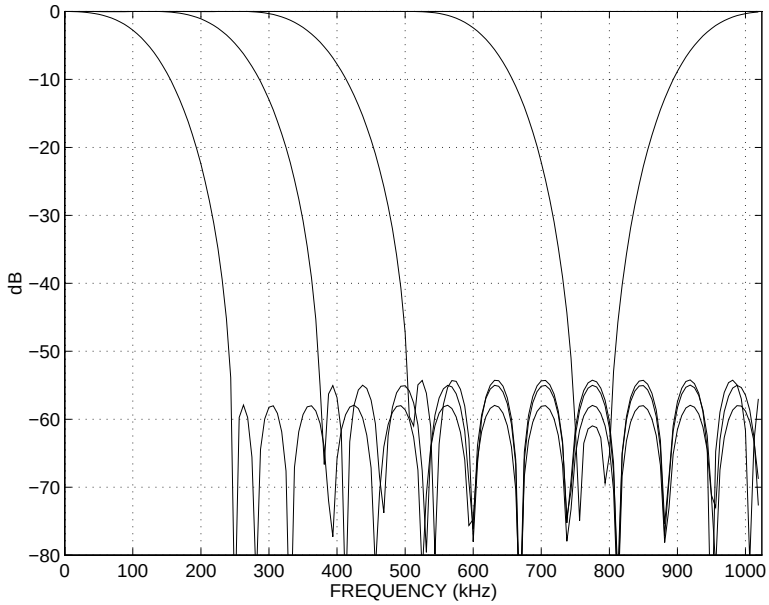


Figure 9.12: Prototype transfer functions $|H_p(f)|, t = 0.9, L_1 = 20$ for complex $\Delta\Sigma$

in dB for the overall system is the same as that for any particular channel. We are also conducting multi-channel simulations, the results of which we will describe in [83], to verify this.

9.8.2 Empirical Comparison With Time Sliced System

Let us finally make a limited comparison of practical performance with respect to implementation imperfection (not absolute performance) between two-channel multi-band and time sliced systems used to convert low frequency signals. The practical performance results for the single channel band-pass modulators described in Chapter 7 suggest that the advantage of using second order complex NTFs over fourth order real NTFs will also be lost in multi-band $\Delta\Sigma$ systems. Consequently, a multi-band system using fourth order real modulators is considered in the comparison with the time sliced system using second order modulators. In [84] the authors, who have proposed the time sliced method, have shown that the SNR obtained from of a low-pass second order $\Delta\Sigma$ modulator is improved by 12 dB using two channels. Although a nice mechanism to alleviate SNR degradation, due to aliasing, is described in [84], the full 15 dB improvement was not obtained due to residual aliasing effects resulting from 0.1% capacitor mismatch. As shown in [84], finite opamp

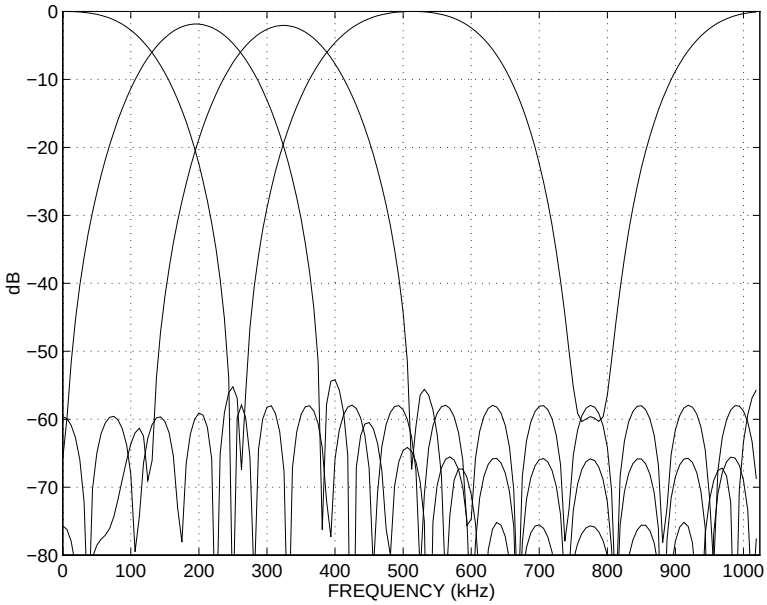


Figure 9.13: Filter bank transfer Functions $|G_p(f)|, t = 0.9, L_1 = 20$ for complex $\Delta\Sigma$

gain problems are actually alleviated with more channels although the effects of opamp gain mismatch between channels are not explicitly discussed. The oversampling ratio of the signal was $R = 100$. Let us consider a similar set of parameters using results obtained for $\omega_c = 0.2$ (a relatively low frequency location) and $R = 128$, which implies $R_B = 256$ for $P = 2$ channels instead of $R_B = 128$ with only $P = 1$ channel. To obtain the SNR degradation For $R_B = 256$, we interpolate results between $R_B = 128$ and 512 based on Fig 7.10(c) and results for $R_B = 512$ (not shown). Doing so under the condition of 60 dB finite opamp gains which were mismatched, the *additional* SNR degradation by going to two channels is about 1 dB. The *additional* SNR degradation with two channels with 0.1% capacitor ratio mismatch is less than 1 dB for a total SNR degradation of less than 2 dB for the multi-band system. Consequently, only about 10 dB improvement in SNR would result from increasing the number of channels to 2 instead of the expected 12 dB. For this comparison we observe that the frequency slicing system using real fourth order modulators was able to obtain 10 of the 12 dB expected improvement in SNR whereas the time slicing system was able to obtain about 12 of the 15 dB expected improvement in SNR.

Chapter 10

Conclusions

10.1 Summary of Results and Contributions

We have examined PCM and delta-sigma A/D conversion from a general point of view of frequency and time slicing. Frequency slicing involves quantizing different portions of the signal frequency spectrum using several quantizers in parallel and then combining the results of the quantizers to form an overall result. Time slicing involves quantizing various groups of time domain signal samples with a number of quantizers in parallel and then combining the results of the quantizers to form an overall output. The performance of such systems under ideal conditions has been compared. Our analysis of PCM systems reveals that for a given signal band and sampling frequency at which the quantizers operate, increasing the number of channels leads to no improvements in resolution. On the other hand, a time interleaved system leads to an improvement in the resolution by 3 dB for every doubling of the number of channels in the system.

Time slicing for delta-sigma systems has been extended to include band-pass modulators. For the case of $\Delta\Sigma$ converters, the SNR of a time sliced system employing delta-sigma modulators with order L lowpass, order L complex band-pass, or order $2L$ real band-pass noise transfer functions (NTFs) improves at a rate of $(6L + 3)$ dB per octave increment in the number of modulators. However, the SNR of frequency sliced systems employing such modulators improves at a rate of $(6L)$ dB per octave increment in the number of modulators. The source of this difference in performance between the frequency and time

slicing approaches has been discussed. Within the no overload framework assumed, the difference results from the fact that in the frequency sliced system the amplitude ranges which need to be converted by each quantizer is, in general, not reduced by a factor of $\sqrt{P}$ where P is the number of channels.

A frequency sliced direct multi-band architecture has been proposed for $\Delta\Sigma$ converters. The architecture uses modulators which realize different NTFs for different portions of the signal band. Each band (channel) is converted in parallel. A bank of FIR filters attenuates the out-of-band noise for each band (which in general need not be of equal size) and can easily achieve perfect reconstruction of the signal component. A design procedure has been provided for the design of the filter bank with reduced computational complexity.

The use of complex band-pass $\Delta\Sigma$ NTFs in delta-sigma systems has been proposed and quantitatively compared with the use of real NTFs. Performance comparisons include the presence of implementation imperfections such as finite opamp gain and capacitor ratio mismatch for several values of oversampling ratio and center frequency. Linearized analysis revealed that complex second order NTFs performed much better than real fourth order NTFs for all center frequencies when stability requirements of the fourth order NTFs were taken into account. However, behavioral simulations show that this advantage is lost in the presence of implementation imperfections. Behavioral simulations which modeled implementation imperfections such as integrator leakage due to finite opamp gain and capacitor ratio mismatch were performed to quantify performance degradation with these imperfections. These simulations show that at larger oversampling ratios and center frequencies, capacitor matching requirements as well as opamp gain matching requirements limit the performance of complex modulators. Consequently, for a multi-band frequency sliced architecture requiring relatively large oversampling ratios, $\Delta\Sigma$ modulators employing real NTFs are preferable and hold promise for the conversion of wider band signals with more resolution. We now discuss some issues that are of interest for future research.

10.2 Directions for Future Research

The filter bank design procedure can accomodate the case where the signal is a band-pass signal. This signal will eventually need to be demodulated to baseband. Hardware complexity issues related to the demodulation have not been treated in the thesis and are a topic of further investigation.

In the case of higher order NTFs, the NTFs and hence the signal transfer functions (STFs) will in general not be FIR and so the STF will not necessarily be a simple delay but may possess in-band linear phase. The filter bank design procedure which has been presented accomodates the perfect reconstruction of the signal only if the modulator STFs are simple delays. With the use of higher order IIR NTFs and hence STFs, the design of the STFs and the filter bank will need to be addressed jointly.

We have made a limited comparison between the frequency sliced multi-band $\Delta\Sigma$ architecture and a time sliced $\Delta\Sigma$ architecture for the case of 1 and 2 channels. Although the researchers who have proposed the time sliced architecture have examined the performance degradation of the architecture with implementation imperfections for the case of 2 channels, the performance degradation cannot be directly extrapolated to more channels due to all the channels being mutually cross-coupled. This is in contrast to multi-band case where the performance degradation of one channel is directly related to the overall system performance degradation provided the white noise assumptions about the quantizer error still remain valid. Also, the clocking scheme for the time interleaved systems becomes multi phased and hence more complicated as the number of channels increases and the delay chains at the input of the structure (sampled and hold circuits in a switched capacitor implementation) will need to operate at a higher speeds for a given rate the quantizers operate at. Consequently, despite the theoretical advantage of the time interleaved system over the frequency sliced system, further research is required to make a more reasonable comparison when more than 2 channels are considered. Also even though time sharing was not directly applicable to complex band-pass modulators, there may exist other circuit techniques which could restore some of the practical performance loss encountered by the complex modulator. In a practical implementation, the issue of clock jitter for the two types of systems also needs to be explored. A simple analysis of clock jitter provided in

Appendix D shows that the if clock jitter is modeled as a white noise process, the effect on the modulator is to have an additive white noise source at the input of the modulator where this noise source experiences no benefit from noise shaping. Consequently, the variance of the noise source should be much less than the conversion resolution. The frequency sliced system has a simpler clocking scheme than the time sliced system and so may suffer from less clock jitter non-idealities. However, it is reasonable to make accurate comparisons only when detailed implementations of the clocking generation circuits are available [85].

We have extended the use of time slicing to band-pass modulators and have confirmed its validity via simulations performed on a two channel modulator employing second order real band-pass NTFs. Linearized noise sources were used for this purpose in place of the quantizers in the modulators. Preliminary simulations of the time interleaved band-pass modulator with the quantizers indicated that it does not attain its theoretical performance. Consequently, a more thorough investigation of time slicing for band-pass systems under practical conditions is of interest.

Finally, we have shown that, within the no overload framework of analysis, the time interleaved system has a theoretical advantage over the frequency sliced system only for the frequency sliced architectures we have examined – we have not *proven* that no frequency sliced delta-sigma architecture exists whose resolution can improve at a rate of $(6L + 3)$ dB per octave increment in the number of channels as opposed to $(6L)$ dB per octave. Therefore, the search for such an architecture is still an open problem.

Appendix A

Time Interleaved Quantization Analysis

Considering only quantization noise components and writing a Z domain equation, the output noise process can be written as,

$$Y(z) = \sum_{r=0}^{P-1} E_r(z^P) z^{-(P-1-r)} \tag{A.1}$$

In the time domain, we have, for each individual channel,

$$y_r[n] = \begin{cases} e_r\left[\frac{n}{P}\right] & n \bmod P = 0 \\ 0 & \text{otherwise} \end{cases} \tag{A.2}$$

Note that by itself, $y_r[n]$ is **not** W.S.S. even if $e_r[n]$ is W.S.S. This is because $\mathcal{E}\{y_r[n]y_r[n+k]\}$ $= \mathcal{E}\{e_r[\frac{n+k}{P}]e_r[\frac{n+k}{P}]\}$ when $n \bmod P = 0$ and 0 otherwise i.e. the expectation depends on n.

Let us now consider, $y[n]$.

$$y[n] = \sum_{r=0}^{P-1} y_r[n - (P-1-r)] \tag{A.3}$$

$$y[n] = \begin{cases} \sum_{r=0}^{P-1} e_r\left[\frac{n-(P-1-r)}{P}\right] & n - (P-1-r) \bmod P = 0 \\ 0 & \text{otherwise} \end{cases} \tag{A.4}$$

What is $\mathcal{E}\{y[n]y[n+k]\}$? where $\mathcal{E}$ is the expectation operator. $y[n]$ is W.S.S. only if $\mathcal{E}\{y[n]y[n+k]\}$ depends on k only and not n.

$$\mathcal{E}\{y[n]y[n+k]\} = \mathcal{E}\left\{\sum_{r=0}^{P-1} e_r\left[\frac{n-(P-1-r)}{P}\right]\sum_{s=0}^{P-1} e_s\left[\frac{n+k-(P-1-s)}{P}\right]\right\} \tag{A.5}$$

Interchanging expectation and summations,

$$\mathcal{E}\{y[n]y[n+k]\} = \sum_{r=0}^{P-1}\sum_{s=0}^{P-1}\mathcal{E}\left\{e_r\left[\frac{n-(P-1-r)}{P}\right]e_s\left[\frac{n+k-(P-1-s)}{P}\right]\right\} \tag{A.6}$$

If we assume that $e_r[n]$ and $e_s[n]$ are uncorrelated for $r \neq s$, then

$$\mathcal{E}\{y[n]y[n+k]\} = \sum_{r=0}^{P-1}\mathcal{E}\left\{e_r\left[\frac{n-(P-1-r)}{P}\right]e_r\left[\frac{n+k-(P-1-r)}{P}\right]\right\} \tag{A.7}$$

For $\mathcal{E}\{y[n]y[n+k]\}$ to not depend on n, we *simultaneously* need **for all** n

1. $[n-(P-1-r)] \bmod P = 0$
2. $[n+k-(P-1-r)] \bmod P = 0$

If these two conditions are satisfied then, $e_r\left[\frac{n-(P-1-r)}{P}\right]$ and $e_r\left[\frac{n+k-(P-1-r)}{P}\right]$ are never zero for any value of n and we will see that the expectation does not depend on n. Clearly, condition 1. is satisfied only if we choose $r = P-p$ where $p = [n-(P-1)] \bmod P$. This is always possible because this value of r satisfies $0 \leq r \leq P-1$. Given this choice of r, condition 2. is satisfied only if we have $k \bmod P = 0$. Thus, we have

$$\mathcal{E}\{y[n]y[n+k]\} = \begin{cases} \mathcal{E}\left\{e_r\left[\frac{n-(P-1-r)}{P}\right]e_r\left[\frac{n+k-(P-1-r)}{P}\right]\right\} & k \bmod P = 0 \\ 0 & \text{otherwise} \end{cases} \tag{A.8}$$

The summation over r is gone because only one value of r allows $r \bmod P = 0$. Assuming that $e_r[n]$ are W.S.S., then by definition, $\mathcal{E}\{e_r[a]e_r[b]\} = R_{e_r}[b-a]$ where R_{e_r} is the autocorrelation function of $e_r[n]$. Then,

$$\mathcal{E}\{y[n]y[n+k]\} = \begin{cases} R_{e_r}\left[\frac{k}{P}\right] & k \bmod P = 0 \\ 0 & \text{otherwise} \end{cases} \tag{A.9}$$

Although for any n we can pick r such that there is no n dependence in the argument of R_{e_r}, the r in the subscript still depends on n. However, if we further assume that $R_{e_r} = R_e \forall r$,

then we have,

$$\mathcal{E}\{y[n]y[n+k]\} = \begin{cases} R_e\left[\frac{k}{P}\right] & k \bmod P = 0 \\ 0 & \text{otherwise} \end{cases} \tag{A.10}$$

Consequently, given the assumption that all e_r are mutually uncorrelated and the assumption that they all have the same autocorrelation function, we showed that $\mathcal{E}\{y[n]y[n+k]\}$ depends only on k. Therefore, $y[n]$ is W.S.S. and so we can write,

$$R_y[k] = \begin{cases} R_e\left[\frac{k}{P}\right] & k \bmod P = 0 \\ 0 & \text{otherwise} \end{cases} \tag{A.11}$$

Equation (A.11) implies

- $\text{var}(y[n]) = R_y[0] = R_e[0] = \text{var}(e[n])$

  This result is stated but not proven in [59]. This has also been verified via simulation.

- $S_y(e^{j\omega}) = S_e(e^{j\omega P})$

where $S_y(e^{j\omega})$ and $S_e(e^{j\omega})$ are the power spectral densities of $y[n]$ and $e[n]$ respectively.

As an aside unrelated discussion let us quickly examine the result downsampling a W.S.S. signal has on its stationarity. The downsampling operation can be represented as

$$y[n] = x[nM] \tag{A.12}$$

where $x[n]$ is a W.S.S. process i.e. $\mathcal{E}\{x[n]x[n+k]\} = R_x[k]$ does not depend on n. We find that $\mathcal{E}\{y[n]y[n+k]\} = \mathcal{E}\{x[nM]x[(n+k)M]\} = \mathcal{E}\{x[nM]x[nM+kM]\} = \mathcal{E}\{x[kM]\}$ where the last step is obtained from the W.S.S. property of $x[n]$. Therefore, $y[n]$ is W.S.S. since $\mathcal{E}\{y[n]y[n+k]\}$ depends only on k. Consequently we can write,

$$R_y[k] = R_x[kM] \tag{A.13}$$

Appendix B

SNR Calculation Techniques

B.1 Frequency Domain SNR Calculation Techniques

This appendix discusses techniques for measuring the signal to noise ratio (SNR) at the output of delta-sigma converters. In a simulation environment, the SNR can be measured in both the frequency or time domains whereas in a laboratory environment, only frequency domain measurements are reasonable. For the simulation environment, the frequency domain technique is more popular and does not require explicit filtering of the modulator output. Consequently, most of this appendix is spent on the frequency domain techniques. A conventional and popular method for measuring the performance of A/D converters involves using a pure tone (sinusoid) as an analog input, measuring the digital output codes over a large number of samples, computing the spectrum of the output and then computing the signal to quantization noise ratio in the frequency domain. If we neglect any noise due to circuit non-idealities, then the output spectrum consists of the signal at the signal frequency as well as quantization noise at the signal frequency and other frequencies. Since it is not obvious how to separate the quantization noise (henceforth referred to as just noise) from the signal at the signal frequency, the noise at the signal frequency is ignored and all the power at this frequency is considered to be signal power. Thus the SNR is the output power at the signal frequency divided by the sum of the output power at all other frequencies.

In practice, the spectrum is computed by taking a N point Discrete Fourier Transform

(DFT) of the output using N output samples where the DFT is a N point frequency sampled version of the Discrete Time Fourier Transform (DTFT). Since the DFT is usually computed using a Fast Fourier Transform (FFT), N is a power of 2. Henceforth the terms FFT and DFT will be used interchangeably. Also, recall that except for the DFT bins 0 and $\frac{N}{2}$, the DFT of a real (not complex) time sequence is symmetric.

Let $y[n]$ be a signal containing a sinusoid at frequency f_I and noise at all other frequencies. If $Y(k)$ is the FFT and $k_I \neq 0, \frac{N}{2}$ is the DFT bin corresponding with the frequency f_I, then the SNR is defined as

$$SNR = \frac{2|Y(k_I)|^2}{2\sum_{k=k_l}^{k_u}|Y(k)|^2}$$

where k_l and k_u are the DFT freqeuncy bin corresponding to with the lower and upper band edge frequency f_u and f_l over which the SNR is to be measured. Note that this definition of SNR does not consider any quantization noise outside of the signal band assuming that an ideal filter can filter out any noise outside the band of interest. If one filtered $y[n]$ with an approximation to an ideal filter to produce the filtered signal $w[n]$, the final SNR which would then reflect the effectiveness of the filter would be defined as,

$$SNR = \frac{2|W(k_I)|^2}{|W(0)|^2 + |W(\frac{N}{2})|^2 + 2\sum_{k=0,k\neq 0,k\neq\frac{N}{2}}^{\frac{N}{2}-1}|W(k)|^2}$$

Since explicit filtering of the signal is assumed to occur, the noise is now measured over all the DFT bins except the sinusoid signal bin.

B.1.1 A Practical Problem: Spectral Leakage

Forgetting about the presence of quantization noise temporarily, let us consider a practical problem associated with the computation of the spectrum of a pure sinusoid from a finite number of samples. The problem is called *spectral leakage* or *frequency smearing* and occurs because by taking only N samples of the sinusoid in the time domain, we have windowed the sinusoid with a rectangular window. This window function, $r[n]$, takes on a value of 1 over the duration of the N sinusoid samples which are taken and is zero everywhere else. By windowing the sinusoid in the time domain, we have in the frequency domain convolved the DTFT spectrum of the sinusoid, which is an impulse, with the DTFT spectrum of the

rectangular window (also known as the Dirichlet window) which is a *sinc* (also known as the Dirichlet Kernel).

The *sinc* spectrum consists of a mainlobe, attenuated sidelobes, and spectral nulls in between adjacent lobes. Convolving the impulse with the *sinc* produces a frequency smearing of the sinusoid, producing a mainlobe representing the smeared signal due to convolution of the impulse with the *sinc* mainlobe. Also produced are sidelobes due to convolution of the impulse with the *sinc* sidelobes. The windowed spectrum also contains spectral nulls in between adjacent lobes.

By taking a N point FFT, this windowed DTFT spectrum is sampled at N equally spaced frequency samples.

$$\omega = \frac{2\pi}{N}k, \qquad k = 0 \ldots N-1$$

From this we have the FFT index or bin number k in terms of the real frequency variable f as,

$$k = \frac{f}{f_s}N$$

Numerically, the sinusoid signal bin occurs at

$$k_I = \frac{f_I}{f_s}N$$

If the value k_I computed as above is an integer, the sinusoid frequency f_I falls on a FFT bin. This occurs when N represents an *integer number of periods* of the sinusoid. This can be seen as follows. Consider a sinusoid sampled at a frequency $f_s = Mf_I$ such that a total of N samples are collected. The number of samples collected per period is M. So we have,

$$\#periods = \frac{\#periods}{sample} \times \#samples$$

$$\#periods = \frac{1}{M}N$$

$$\#periods = \frac{f_I}{f_s}N$$

But this is the same as the quantity k_I which represents the calculated signal frequency in the FFT domain. So if the number of periods collected is an integer, the signal frequency

falls on a FFT bin. Let us call this condition *coherent sampling.* One can also show that requiring N to represent an integer number of periods of the sinusoid forces the spectral nulls of of the windowed spectrum to fall on FFT bins. In fact, when N represents an integer number of periods of the sinusoid, the signal frequency f_I and the nulls in the windowed spectrum are the *only* frequencies which fall on FFT bins and the effects of frequency smearing are not *observable* in the FFT spectrum. Let us verify this analytically.

Consider a discrete time sinusoid $y[n]$ which was sampled from a continuous time sinuosid,

$$y[n] = Asin(\omega_I n)$$

Expanding in terms of complex exponentials we have,

$$y[n] = \frac{A}{2j}e^{j\omega_I n} - \frac{A}{2j}e^{-j\omega_I n}$$

Now let the sinusoid be windowed by the rectangular window $r[n]$ to produce the windowed signal $v[n]$

$$r[n] = \begin{cases} 1 & 0 \leq N-1 \\ 0 & \text{otherwise} \end{cases}$$

$$v[n] = y[n]r[n]$$

$$v[n] = \frac{A}{2j}e^{j\omega_I n}r[n] - \frac{A}{2j}e^{-j\omega_I n}r[n]$$

The DTFT spectrum of the windowed signal is then

$$V(e^{j\omega}) = \frac{A}{2j}R(e^{j(\omega-\omega_I)}) - \frac{A}{2j}R(e^{j(\omega+\omega_I)})$$

The spectrum of the window is a $sinc$:

$$R(e^{j\omega}) = \frac{sin(\frac{\omega N}{2})}{sin(\frac{\omega}{2})}e^{-j\omega\frac{(N-1)}{2}}$$

So the DTFT spectrum of the rectangular windowed signal becomes,

$$V(e^{j\omega}) = \frac{A}{2j}\frac{sin(\frac{(\omega-\omega_I)N}{2})}{sin(\frac{(\omega-\omega_I)}{2})}e^{-j(\omega-\omega_I)\frac{(N-1)}{2}} - \frac{A}{2j}\frac{sin(\frac{(\omega+\omega_I)N}{2})}{sin(\frac{(\omega+\omega_I)}{2})}e^{-j(\omega+\omega_I)\frac{(N-1)}{2}}$$

Plugging in $\omega = \frac{2\pi}{N}k$ and $\omega_I = 2\pi\frac{f_I}{f_s}$ and simplifying we obtain the spectrum $V(k)$ in the frequency sampled DFT domain,

$$V(k) = \frac{A}{2j}\frac{sin(\pi(k - \frac{f_I}{f_s}N))}{sin(\frac{\pi}{N}(k - \frac{f_I}{f_s}N))}e^{-j\pi(1-\frac{1}{N})(k-\frac{f_I}{f_s}N)} - \frac{A}{2j}\frac{sin(\pi(k + \frac{f_I}{f_s}N))}{sin(\frac{\pi}{N}(k + \frac{f_I}{f_s}N))}e^{-j\pi(1-\frac{1}{N})(k+\frac{f_I}{f_s}N)}$$

The magnitude DFT spectrum is then,

$$|V(k)| = \frac{A}{2} \mid \frac{sin(\pi(k - \frac{f_I}{f_s}N))}{sin(\frac{\pi}{N}(k - \frac{f_I}{f_s}N))}e^{-j\pi(1-\frac{1}{N})(k-\frac{f_I}{f_s}N)} - \frac{sin(\pi(k + \frac{f_I}{f_s}N))}{sin(\frac{\pi}{N}(k + \frac{f_I}{f_s}N))}e^{-j\pi(1-\frac{1}{N})(k+\frac{f_I}{f_s}N)} \mid$$

It is clear that if $\frac{f_I}{f_s}N$ is an integer i.e. we have coherent sampling, $|V(k)|$ is zero for all k except for $k = \frac{f_I}{f_s}N$ and $k = -\frac{f_I}{f_s}N$. At these signal bins, $|V(k)| = \frac{A}{2}N$. If we normalize $|V(k)|$ by dividing by N, the value of the magnitude spectrum at the signal bin, $k_I = \frac{f_I}{f_s}N$ is then $\frac{A}{2}$. Recall that the bin $-k_I$ is equivalent to the bin $N - k_I$ and so in the actual output spectrum the two signal bins are k_I and $N - k_I$. Thus the normalized $|V(k)|$ has a value of $\frac{A}{2}$ at bin $N - k_I$.

It is important to note that the frequency smearing is caused not by frequency sampling with the FFT but by the windowing. A judicious choice of N as discussed earlier has merely hidden this frequency smearing.

It should also be noted that in the case of coherent sampling, the power of the windowed spectrum at the FFT signal bins k_I and $N-k_I$ corresponds to only the power of the sinusoid represented by the N discrete-time samples, and not any power produced by smearing of the sinusoid spectrum with either the window mainlobe or sidelobes. For our sinusoid, we saw in the case of coherent sampling that the normalized $|V(k)|$ had a value of $\frac{A}{2}$ at the FFT signal bins. Thus the energy at those bins is $\frac{A^2}{4}$ each giving a total signal energy of $\frac{A^2}{2}$. Since the FFT is zero at all other bins, the power of the spectrum is also $\frac{A^2}{2}$ which corresponds to the power of the continuous time sinusoid.

When we don't have coherent sampling, the signal frequency f_I does not fall on a FFT bin and frequencies other than spectral nulls of the windowed spectrum fall on FFT bin frequency samples and so the frequency smearing (which is always present in the DTFT of

a windowed signal) is seen in the FFT.

Another way of looking at spectral leakage is considering what is occurring in the time domain as explained below. The DFT is just a frequency sampled version of the DTFT. In computing the DFT, of a finite length signal of length N, there is a periodicity which is created in the time domain by a replication of the finite length time signal (this is analogous to the periodicity created in the discrete time frequency domain as a result of the sampling in time of a continuous time signal) [5, pages 527-530]. The DFT then represents the spectrum of the resulting periodic signal. If the finite length signal represents a sinusoid but does not consist of an integer number of periods of the sinusoid then a discontinuity is produced between the original finite length sinusoid samples and the replicas. Because of the periodicity, these discontinuities are of course present in each period of the periodic time signal which now no longer represents a true sinusoid because of these discontinuities. If the finite length signal did consist of an integer number of periods, no discontinuities would be produced between the finite length signal and the replicas. The resulting periodic signal would look like a true sinusoid hiding the fact that originally, the sinusoid was windowed with a finite duration window.

B.1.2 Spectral Leakage Of A Sinusoid: An Example

Let us illustrate what has been discussed so far with a concrete example. The first plot of Fig B.1 shows a pure sine wave of frequency f_I=16 Hz and amplitude 1 unit sampled at f_s=2048 Hz. The number of samples taken and plotted is N=512. The time samples start at n=0. Notice that in this case, $k_I=\frac{f_I}{f_s}N = 4$ which is an integer. Also, note in the context of the last paragraph, that if the sinusoid of frequncy 16 Hz shown in the first plot were replicated (repeatedly) starting after the last sample plotted, the replica of the signal, which starts with a sample height of zero units, would fit neatly resulting in an overall periodic signal with no discontinuities. This is shown in the second plot of Fig B.1 where the 512 samples of the 16 Hz sinusoid has been replicated once. This is possible since N here represents an integer number (4) of periods. The first sample at n=0 corresponds with

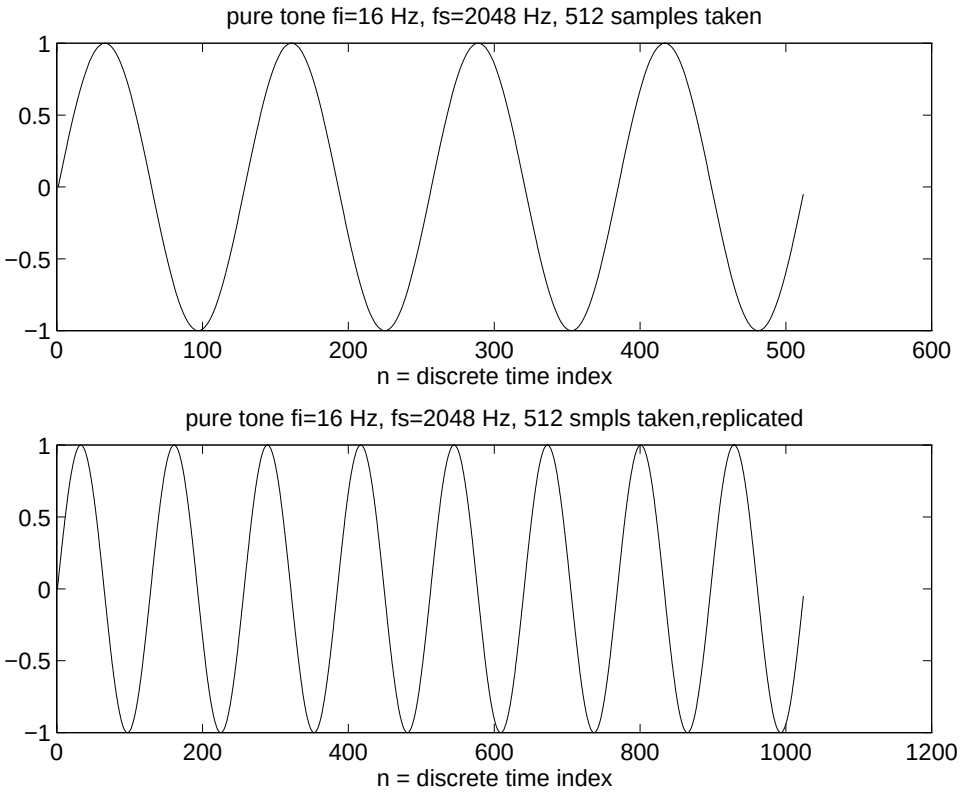


Figure B.1: First plot shows $N = 512$ samples of a sinusoid with frequency f_I=16 Hz. The sinusoid has been sampled with $f_s = 2048$ Hz. Note that the 512 samples consist of an integer number periods of the sinusoid. The second plot shows the first 512 samples being replicated to produce 1024 samples. Note no discontinuity is produced at the boundary between the last sample of the first 512 and the first sample of the repeated 512.

a value of zero units, while the last sample (the 512th) sample corresponds with a negative height (-0.0491) of almost zero units.

The first plot of Fig B.2 shows a pure sine wave of f_I=16.3 Hz and amplitude 1 unit sampled at f_s=2048 Hz. The number of samples taken and plotted is N=512. The time samples start at n=0. Notice that in this case, k_I=$\frac{f_I}{f_s}N = 4.075$ which is not an integer. In the case of the sinusoid of frequency 16.3 Hz shown in the first plot, the 512 samples *do not* correspond to an integer number of periods of that sinusoid. Note that if this sinusoid is replicated (repeatedly) starting after the last sample plotted, there will be a discontinuity between the the 512 samples plotted and the next replica which starts at sample 513. The first of the 512 samples starts at n=0 and corresponds with a height of zero units. The 512th sample corresponds to a height of about 0.409 units. However, the first replica of the signal

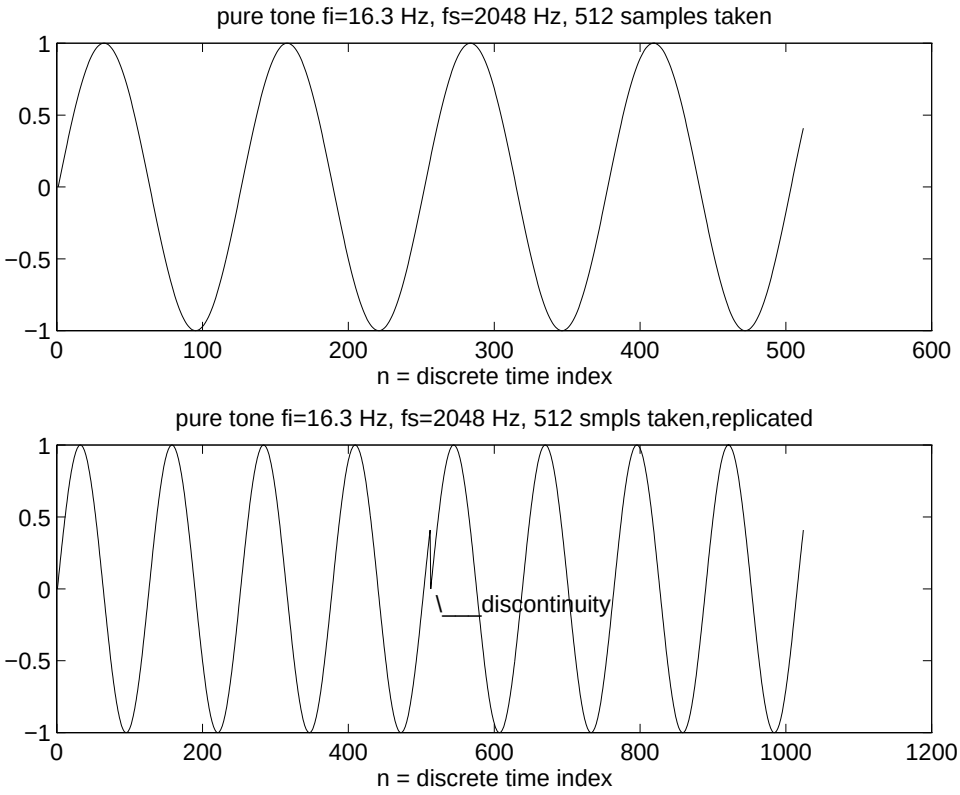


Figure B.2: First plot shows $N = 512$ samples of a sinusoid with frequency f_I=16.3 Hz. The sinusoid has been sampled with $f_s = 2048$ Hz. Note that the 512 samples do not consist of an integer number periods of the sinusoid. The second plot shows the first 512 samples being replicated to produce 1024 samples. Note the discontinuity produced at the boundary between the last sample of the first 512 and the first sample of the repeated 512.

starts at a height of zero units. Thus a discontinuity is produced between the original signal and its replicas. This is illustrated in the second plot of Fig B.2. Let us now go to Fig B.3. The first plot shows a plot of a $N = 4096$ point normalized FFT magnitude (dB) spectrum of the aforementioned 16 Hz sinusoid. The second plot shows this spectrum magnified in the region of interest and plotted discretely so that individual FFT bins are discernible. As expected, we can see only one non-zero component at the FFT bin k=64 which is the same as k_I=64 (for $N = 4096$). In the actual plot, the non-zero value is seen at k=65 because in the MATLAB software used to generate the simulations, $k_{MATLAB} = k+1$. Other values in these plots, due to finite precision computation, are more than 300 dB below the value at k_I and can be ignored. The first plot of Fig B.4 shows a plot of a 4096 point FFT magnitude spectrum of the 16.3 Hz sinusoid. The second plot of this figure shows the region of interest

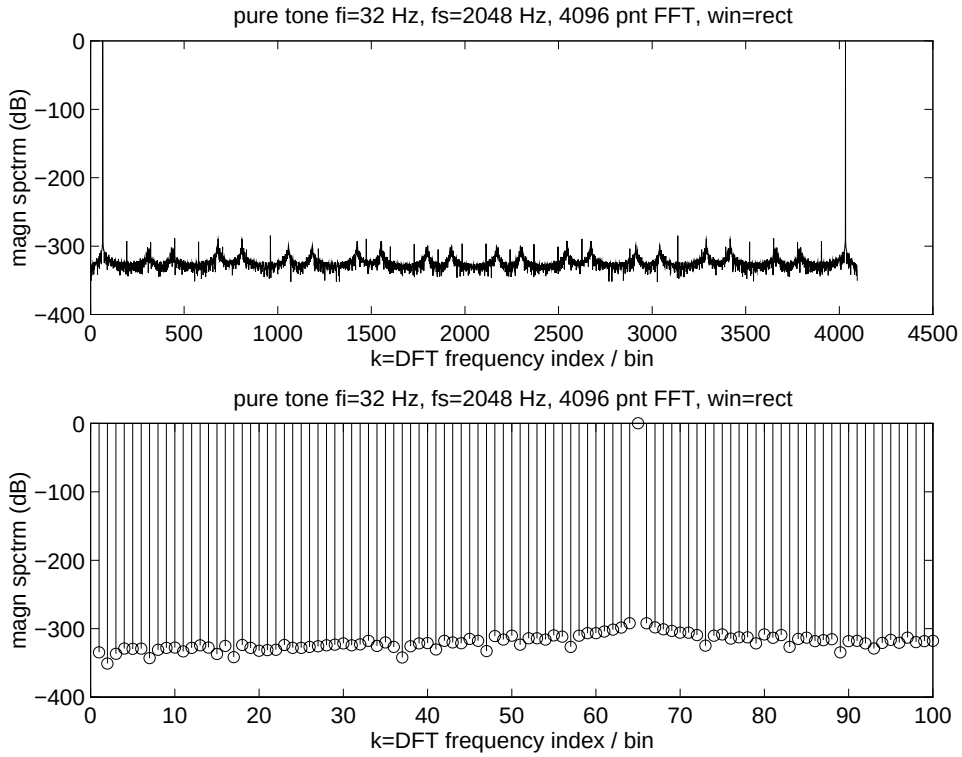


Figure B.3: Spectrum of coherently sampled and rectangularly windowed sinusoid. The second plot is a blow up in the region of interest and is plotted discretely.

magnified and plotted discretely. These plots clearly reveal the effect of spectral leakage since the FFT spectrum now contains many non-zero values. The value k=64 (or 65 in MATLAB) now no longer represents the frequency f_I but some other frequency f which is a smeared version of the signal frequency f_I due to smearing with the mainlobe of the rectangular window. Other non-zero values of the FFT at indices close to k=64 represent energies produced by the smearing of the signal frequency with the window mainlobe while non-zero values of the FFT further away from k=64 represent energies produced by the smearing of the signal frequency with the window sidelobes.

B.1.3 Dealing With Spectral Leakage

In practice, there are three ways to deal with the problem of spectral leakage encountered during the FFT testing of an ADC [86].

1. The first option forces the quantity k_I to be an integer M by setting the ratio $\frac{f_I}{f_s}$ to

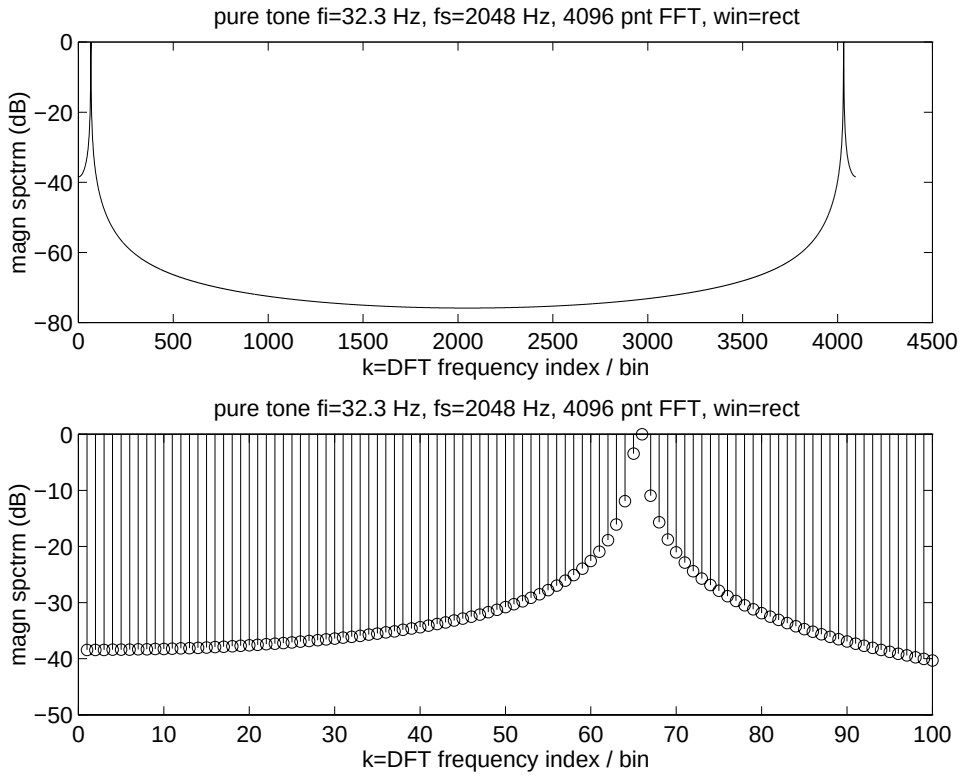


Figure B.4: Spectrum of an in-coherently sampled and rectangularly sinusoid. The second plot is a blow up in the region of interest and is plotted discretely.

be $\frac{M}{N}$ by using frequency divider and phase locked-loop (PLL) hardware.

2. The second option assumes the availability of a precision function generator allowing the control of sine wave frequncies down to fractions of a hertz. The sine wave frequency is chosen such that for a given f_s and N, M again turns out to be an integer.

3. The third option is to alleviate the problems assoicated with spectral leakage by using a window function other than the rectangular window.

If one of the first two options is available, coherent sampling is guaranteed and no spectral leakage is observable in the FFT of the rectangular windowed signal. However, it will be seen that in some situations it will be advantageous to use a window other than the rectangular one even if coherent sampling is possible. All this is discussed in detail below.

B.1.4 Spectral Leakage In The Presence Of Noise & The Use Of Windows

Suppose we have N samples of a signal which represents a quantized sinusoid i.e. a signal containing a sinusoid at frequency f_I and noise at all other frequencies. In this case not only must we try to avoid the type of severe spectral leakage seen in Fig B.4 but we must ensure that the noise is not eclipsed by any window artifacts.

The Case of Coherent Sampling

First consider the case where N represents an integer number of periods of the sinusoid. In that case, k_I is an integer and the signal frequency f_I falls on a FFT bin in the rectangular windowed spectrum. No other FFT bins contain any energy due to smearing of the sinusoid *signal* spectrum with either the window mainlobe or sidelobe. Let us ignore any noise which also occurs on the FFT bin k_I. Each real *noise* frequency is also smeared by the *sinc* and produces a mainlobe as well as sidelobes. As a result, each real frequency of the windowed noise spectrum will have a contribution from the noise at that frequency in the original noise spectrum, contributions due to smearing of adjacent frequencies due to the *sinc* mainlobe, and contributions due to smearing of more distant frequencies due to the *sinc* sidelobes. A set of these noise frequncies will be mapped to FFT bins after a FFT of the windowed noise spectrum is taken. Consequently, at each FFT bin of the windowed spectrum, the energy present is somewhat greater than the energy present at the original noise spectrum.

Depending on what the noise spectrum looks like, the above situation can be satisfactory or deadly. The rectangular window has a sidelobe attenuation of about 13 dB [87]. This means that at a particular frequency, f, there can be noise contribution from the sidelobes of noise at other frequencies. If the noise spectrum is flat, then these sidelobe contributions will be at least 13 dB below the noise at the frequency f. However, suppose the noise spectrum is not flat but has a high pass shape as is the case at the output of a low-pass delta-sigma modulator. In this case, the noise at f will have sidelobe contributions from frequencies greater than f. Since the magnitude of the noise for frequencies greater than f can be much higher than the noise at f, the sidelobe contributions of the higher frequency

Window	**Sidelobe Attenuation (dB)**	*spread*
Rectangular (rect)	13	0
Hamming (hm)	43	2
Blackman (blk)	58	3
Blackman-Harris 3 term (bh3)	67	3
Blackman-Harris 4 term (bh4)	92	4

Table B.1: Summary of window properties: side lobe attenuation in dB and spread in number of bins.

noise (attenuated by only 13 dB) may eclipse the true value of the noise at f. When the output of a delta-sigma modulator is windowed with a rectangular window, this effect may be seen in terms of the baseband noise shaping being destroyed and appearing flat.

Consequently, to be safe, the output of the A/D converter should be windowed by a window with a much higher sidelobe attenuation than that provided by the rectangular window. Table B.1 provides a list of commonly used windows and their sidelobe attenuations. The windows are defined for $n = 0 \ldots N-1$ as follows,

$$hm[n] = 0.5 - 0.46\cos\left(\frac{2\pi n}{N}\right)$$

$$blk[n] = 0.42 - 0.50\cos\left(\frac{2\pi n}{N}\right) + 0.08\cos\left(\frac{4\pi n}{N}\right)$$

$$bh3[n] = 0.42323 - 0.49755\cos\left(\frac{2\pi n}{N}\right) + 0.07922\cos\left(\frac{4\pi n}{N}\right)$$

$$bh4[n] = 0.35875 - 0.48829\cos\left(\frac{2\pi n}{N}\right) + 0.14128\cos\left(\frac{4\pi n}{N}\right) - 0.01168\cos\left(\frac{6\pi n}{N}\right)$$

The DTFT spectrum of these windows also have mainlobes and sidelobes. The sidelobe attenuation values (relative to the peak value of the mainlobes) and the window expressions were obtained from [87]. All of the windows have the advantageous property that when coherent sampling is ensured, only a finite number of signal bins from the signal mainlobe of the windowed signal is *observed* in the FFT spectrum. For the rectangular window, only one bin of the signal mainlobe is observed. For the other windows, an odd number of

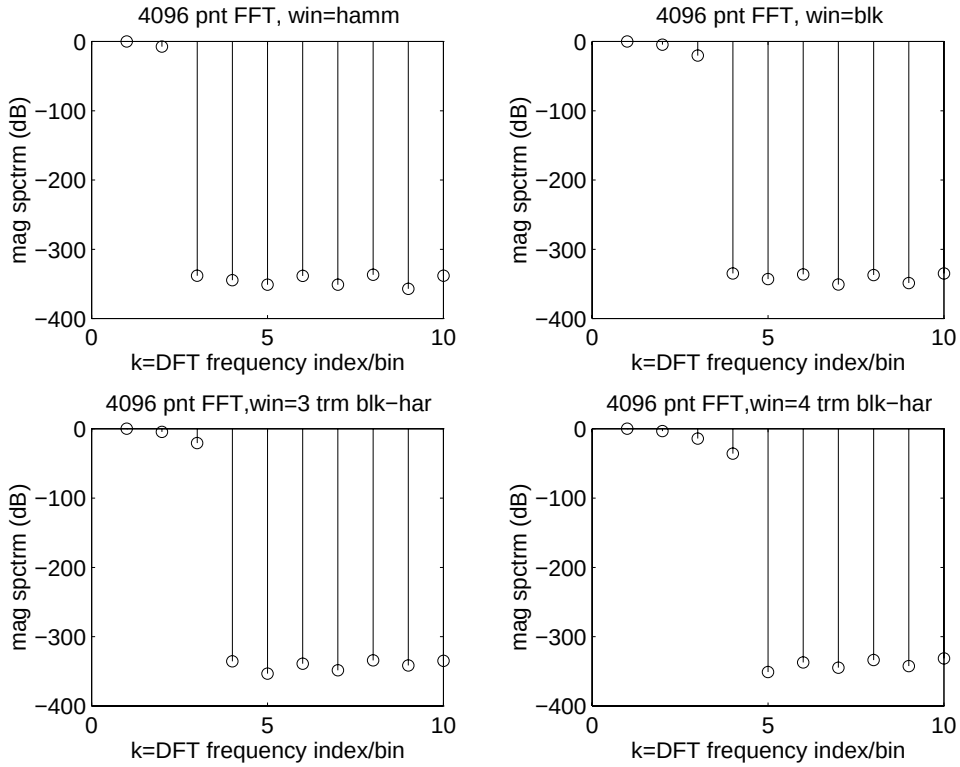


Figure B.5: Spectra of several windows

bins is always observed and the signal is said to be spread by the window. The number of signal bins observed are the bin where the signal spectrum is maximum and also a few bins on each side of this maximum. The number of bins observed to each side of the maximum is called the window spread. Values of window spread are also provided in Table B.1. The FFT spectra of these windows is shown in the four plots of Fig B.5 for a 4096 point FFT. The plots show part of the window mainlobes only – the maximum value of the window spectrum as well as the spread to one side of it. Note that no sidelobes are observable and bins beyond the spread value are more than 300 dB below the maximum and indicate computational noise. The plots are normalized so that the peak value for all of them is 0 dB.

The result of windowing a coherently sampled sinusoid with these windows is shown in Figs B.6 and B.7. The magnitude spectrum plots shown in these figures are those of a 32 Hz sinusoid sampled at 2048 Hz and windowed with a 4096 point window. All plots are

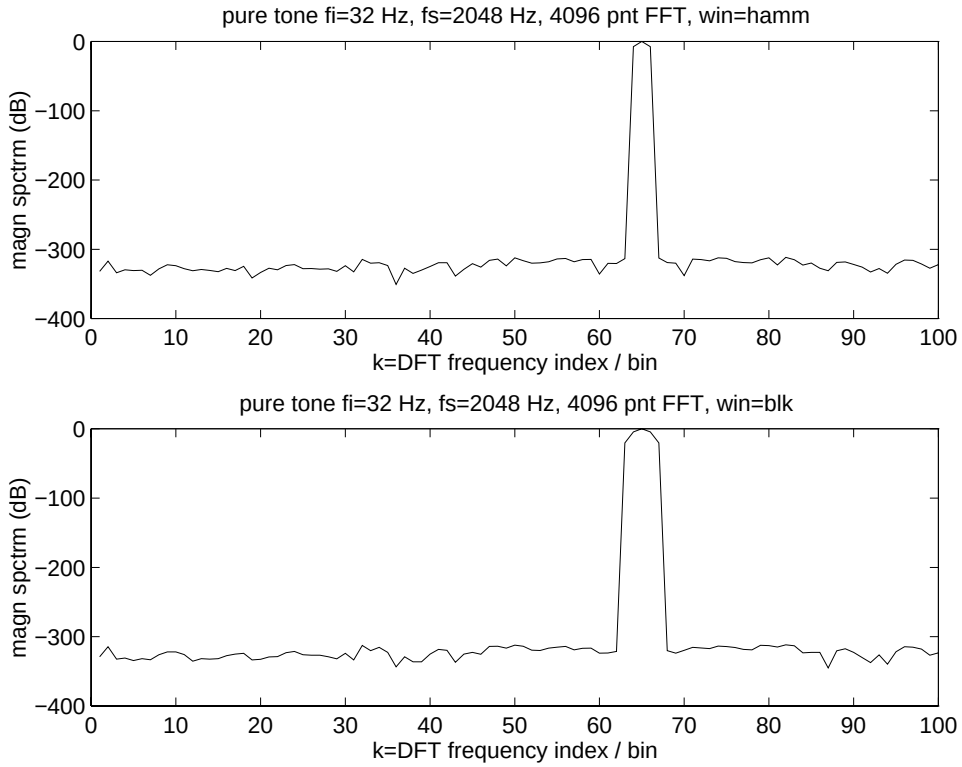


Figure B.6: Spectra of coherently sampled sinusoid windowed with Hamming and Blackman windows. First plot is with Hamming window, second with the Blackman window.

normalized so that the peak value is 0 dB. Note that only a signal mainlobe is observed in the windowed signal – the values of the spectrum at all other frequencies is below 300 dB indicating computational noise.

A few observations regarding the window definitions provided should be made. The definitions of the windows obtained from [87] guaranteed the nice property that for a coherently sampled sinusoid which is windowed, only the mainlobe of the windowed sinusoid is observed in the FFT. These same windows are defined differently elsewhere [5], [88], [52]. These alternative definitions do not have this nice property. It turns out that all of the other window definitions are truly symmetric whereas the definitions provided by Harris in [87] have an implied even symmetry. Implied even symmetry means that the window is not truly symmetric but if it were replicated starting at the next time sample, the window and its replicas would line up neatly resulting in a perfectly periodic time sequence without pro-

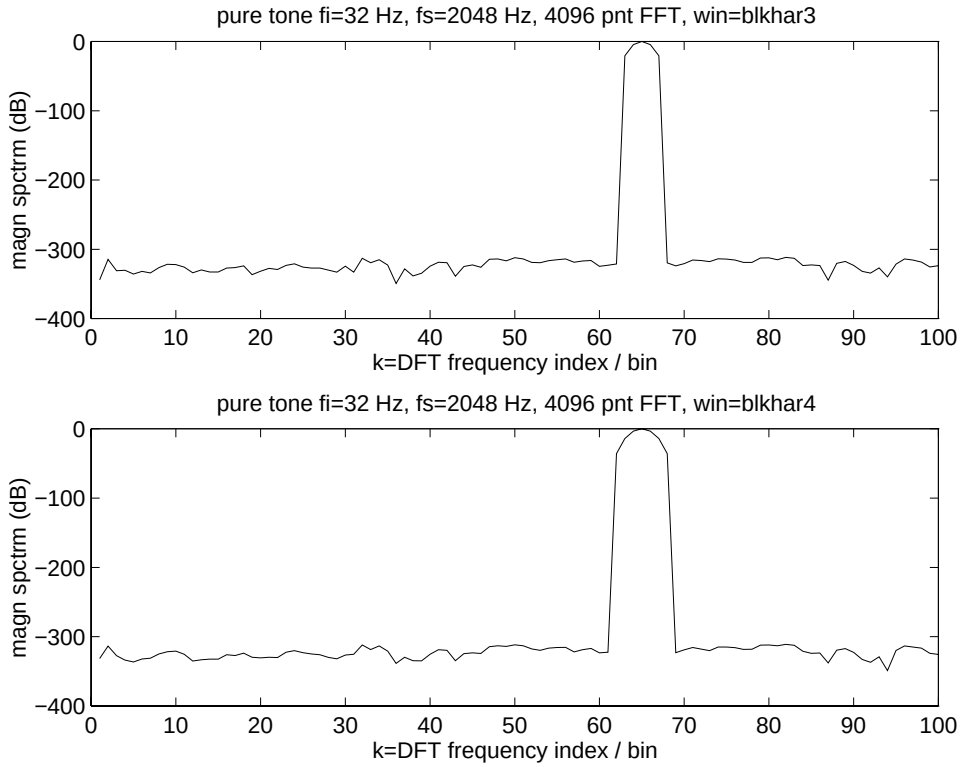


Figure B.7: Spectra of coherently sampled sinusoid windowed with 3-term and 4-term Blackman-Harris windows. First plot is with the 3-term window, second with the 4-term window.

ducing any discontinuities between period boundaries. Note that this property of replicas lining to prevent discontinuities at period boundaries ensured the absence of spectral leakage for the coherently sampled and rectangular windowed sinusoid. Finally, it should also be mentioned that the window definitions provided in [5], [88], and [52] are not necessarily "incorrect" - they are merely inconsistent with the type of spectral analysis we are trying to perform. The alternative window definitions may have little or no adverse consequences in their use in other areas of digital signal processing such as digital filter design based on window methods.

The Case of In-coherent Sampling

Now let us consider the case where N does not represent an integer number of periods of the sinusoid i.e. we do not have coherent sampling. No matter which window we use, the signal will produce a signal mainlobe and signal sidelobes and the noise will also produce noise

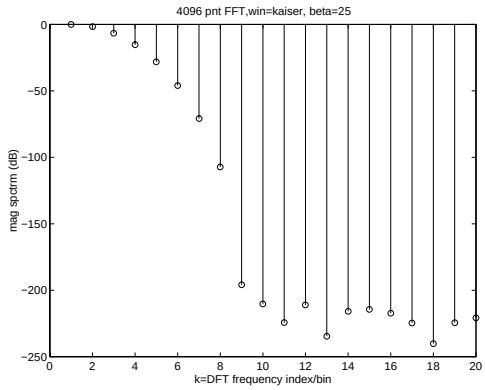


Figure B.8: Spectrum of Kaiser window

mainlobes and noise sidelobes. In this case, the window sidelobe attenuation should again be large so that the true noise at a particular frequency is not eclipsed by noise sidelobes from other frequencies (as was also the criterion in the coherently sampled case). However, in this case the window sidelobe attenuation should also be large enough so that the noise at any frequency is not eclipsed by any contribution from the *signal sidelobe*. Which window is satisfactory depends again on the expected noise spectrum. If the noise spectrum is flat and say the noise floor is 50 dB relative to the signal peak, then using a hamming window is clearly not acceptable since it has a sidelobe attenuation of only 43 dB and so the signal sidelobe can eclipse the noise. In the case of delta-sigma modulation, the noise in the baseband is greatly attenuated relative the signal peak and can have values up to 200 dB below the signal peak. Thus, none of the windows mentioned so far are adequate where coherent sampling is not possible. In this case a kaiser window with a β of 25 is considered. The FFT magnitude spectrum of this window is shown in Fig B.8. Note the spread of 7 and a sidelobe attenuation of about 196 dB.

For the interest of the reader the spectra of an in-coherently sampled sinusoid signal (without noise) which is windowed with the Hamming, Blackman, 3-term Blackman-Harris and 4-term Blackman-Harris windows are shown in Figs B.9 and B.10. The first plot of Fig B.11 shows the Kaiser windowed spectrum ($\beta = 25$) of a 32.3 Hz tone sampled at 2048 Hz with 4096 points. For comparison, the second plot of this figure shows the Kaiser windowed

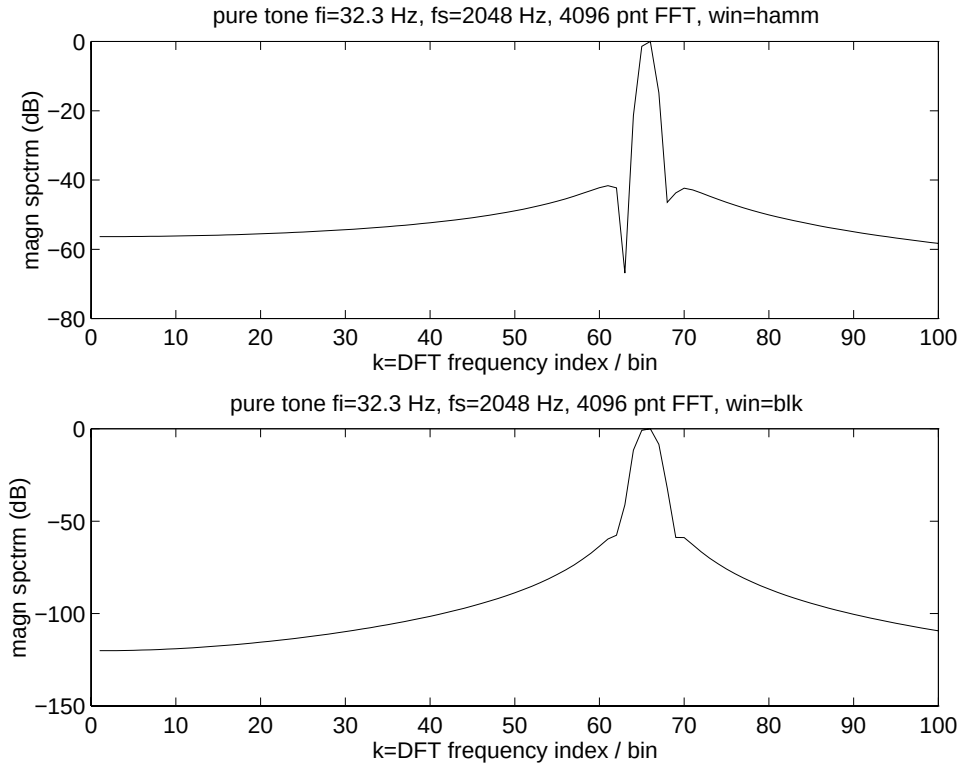


Figure B.9: Spectra of in-coherently sampled, Hamming & Blackman windowed sinusoid

spectrum ($\beta = 25$) of a 32 Hz tone coherently sampled at 2048 Hz with 4096 points. All the plots of Figs B.9 to B.11 are normalized so that the peak signal value is 0 dB. This time note the presence of a signal mainlobe as well as sidelobes with sidelobe attenuations about the same as those in Table B.1 The Kaiser window plots show that, unlike the other windows, the Kaiser window does not have the nice property of not producing any signal sidelobes for a coherently sampled sinusoid. For the 32.3 Hz tone, a signal mainlobe and sidelobe are, of course, present. As mentioned earlier, the window mainlobe for the kaiser window with a $\beta = 25$ consists of 7 bins to each side of the maximum. Incidentally, for the Kaiser window any desired sidelobe attenuation (within numerical precision limits) can be achieved by simply increasing the value of β. The price paid is a wider mainlobe which for a fixed FFT size means that the signal occupies a larger portion of the spectrum, an undesirable effect for an accurate measurement of SNR. In general, most windows have this sidelobe attenuation versus mainlobe width tradeoff. Incidentally, version 3.5 of MATLAB

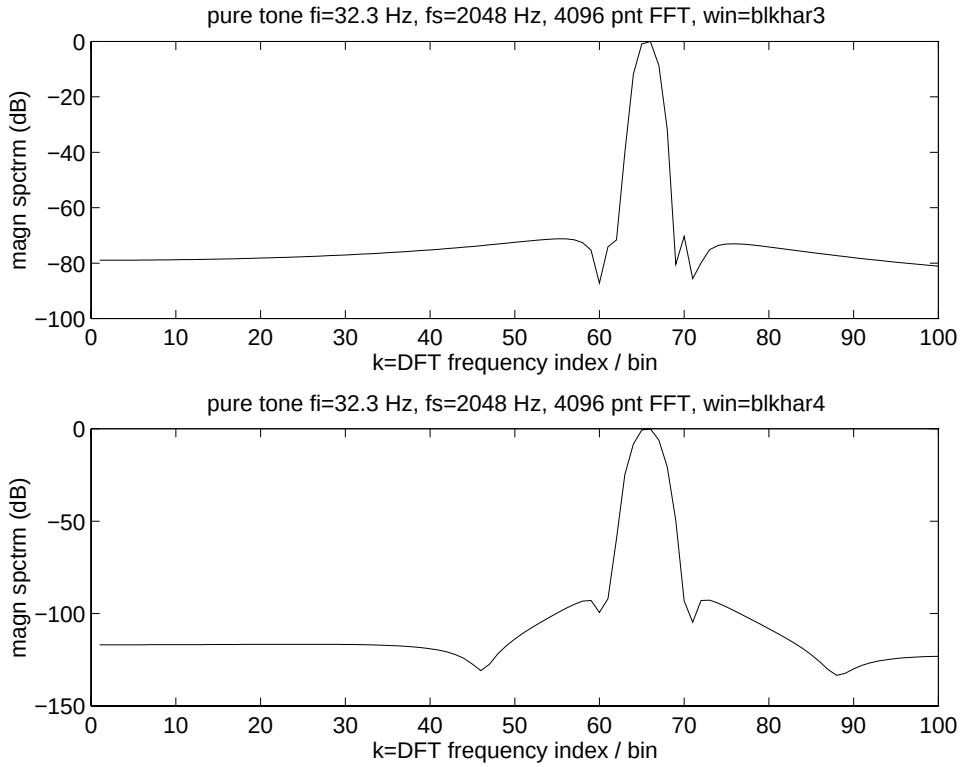


Figure B.10: Spectra of in-coherently sampled sinusoid windowed with 3-term and 4-term Blackman-Harris windows

running on a Sun 4 (32 bit single precision, 64 bit double precision) computer is incapable of producing accurate Kaiser windows for values of β greater than 25.

Finally, let us reiterate the connection between our description of windowing in the frequency domain and what is happening in the time domain. We have seen the effect of windoing a sinusoid with a rectangular window. For the in-coherently sampled sinusoid we observed frequency domain spectral leakage and related this leakage to the time domain by noting that there existed a discontinuity between the time domain sinusoid and its replica. This was shown in the last plot of Fig B.2. Windowing the sinusoid with a window other than the rectangular window allows the sinusoid samples to be smoothly brought to zero at each side instead of abruptly being cutoff as occurs with the rectangular window. The first plot of Fig B.12 redisplays the 512 samples taken from the 16.3 Hz tone sampled at 2048 Hz. The second plot of Fig B.12 shows the time domain plot of a 512 point Blackman window. The

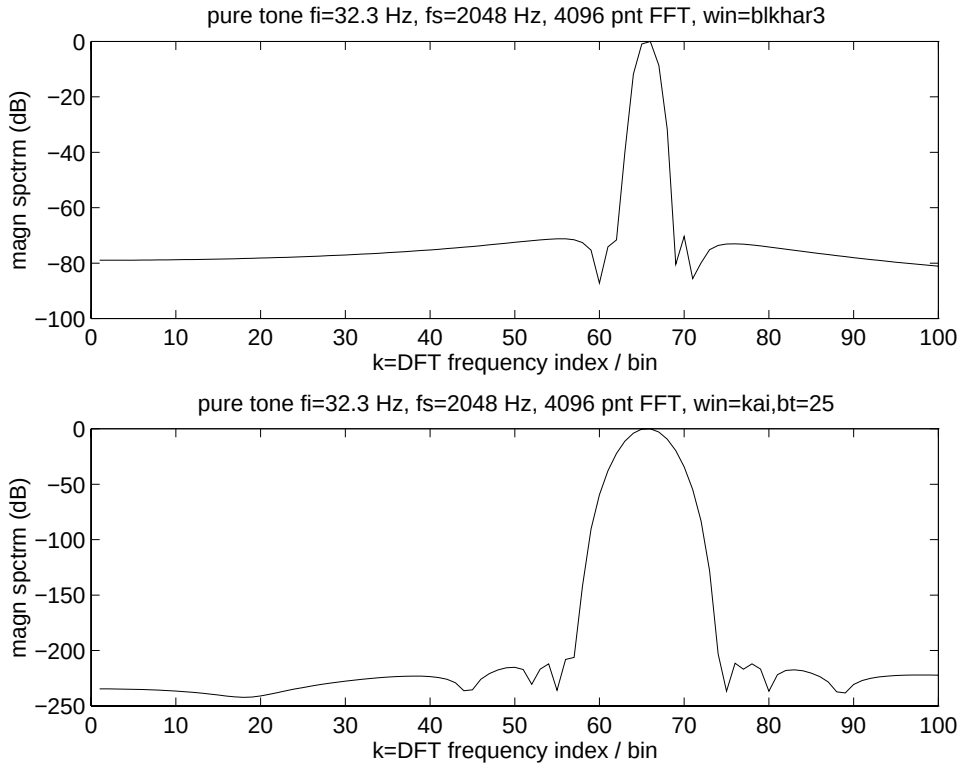


Figure B.11: Spectra of coherently & in-coherently sampled, Kaiser windowed sinusoids

first plot of Fig B.13 shows the windowed version of the tone windowed by the Blackman window. Note that the windowed tone slowly approaches zero at each extreme. The second plot of Fig B.13 shows the windowed tone replicated once. Observe that contrary to what is seen in Fig B.2 (in the case of the rectangular window), the boundary between the first 512 samples and the next 512 samples (the replica) is much smoother showing that the discontinuity has been alleviated. This corresponds with the fact that in the frequency domain, the Blackman window has a much better sidelobe attenuation than the rectangular window. It is interesting to see that the windowed signal visually does not look anything like the original sine wave in the time domain. However, intuitively, it looks like a sum of two or more sinusoids. This makes sense because the windowed time domain plot closely represents the original 16.3 Hz frequency spread into a mainlobe which obviously contains more than one frequency.

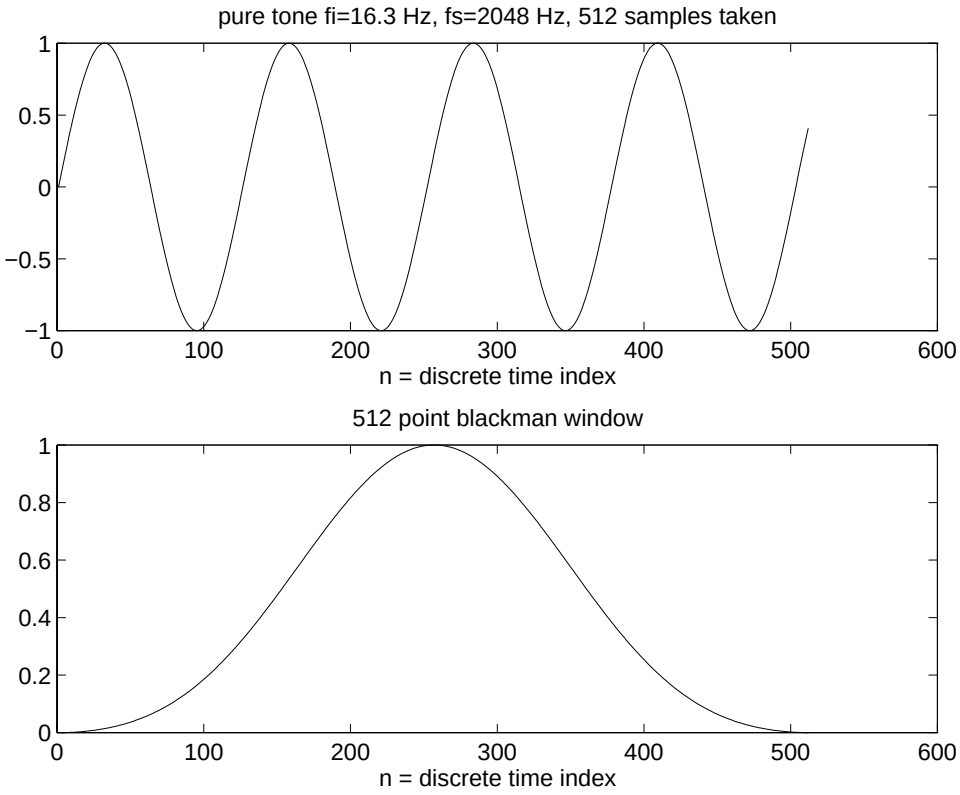


Figure B.12: First plot shows 16.3 Hz sinusoid sampled incoherently with $f_s = 2048$ Hz windowed with rectangular window. Second plot shows the Blackman window.

B.2 Time Domain SNR Calculation Techniques

For computer simulation purposes, there is access to the original analog input data, and the SNR can also be computed in the time domain. If the system is oversampled, we must of course first filter the output of the the A/D converter with a near ideal filter with a given cutoff frequency to attenuate out-of-band quantization noise which should not contribute towards the SNR measurement. The near ideal filter should obviously have a very narrow transition band and an extremely good stop-band attenuation. Let the original signal be $x[n]$, the output of the A/D converter be $y[n]$ and the filtered version of the A/D output be $w[n]$. In this case the error signal is computed as $e[n] = w[n] - x[n]$ and the SNR is defined

$$SNR = \frac{\sum_{n=ni}^{nf} x[n]^2}{\sum_{n=ni}^{nf} e[n]^2}$$

where ni and nf are the initial and final time indices over which the computations are performed. Some issues are important to point out here. First, the filter should have

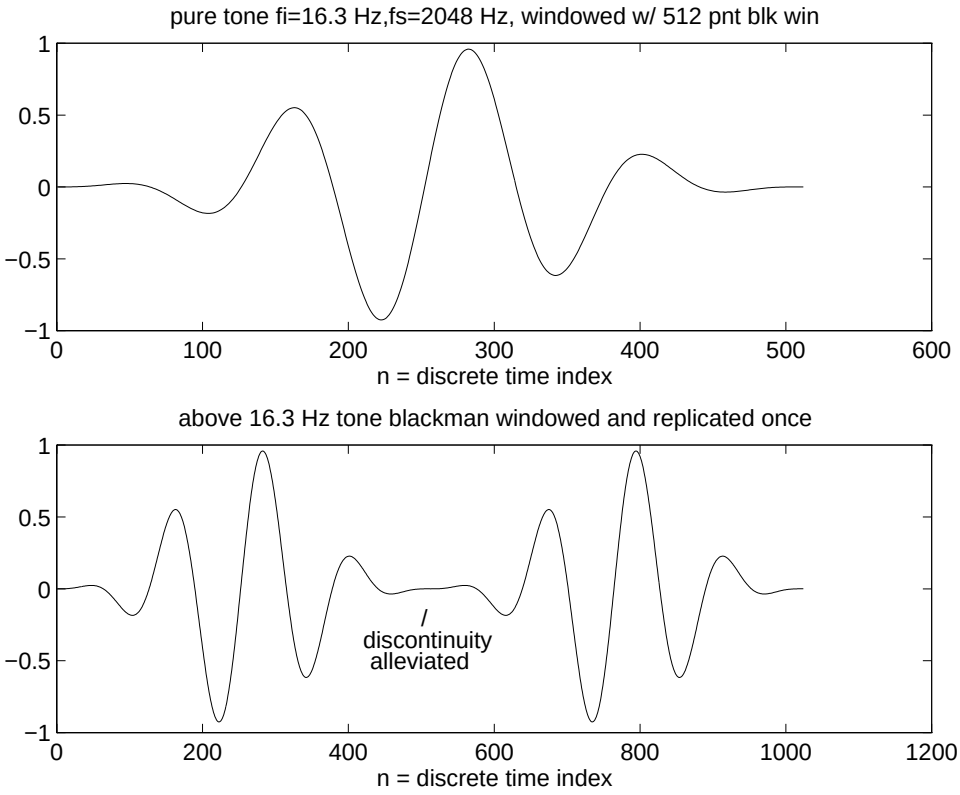


Figure B.13: First plot shows 16.3 Hz sinusoid sampled incoherently with $f_s = 2048$ Hz windowed with the Blackman window. The second plot shows the effect of replicating Blackman windowed sinusoid.

perfectly linear phase i.e. constant group delay such that the delay through the filter should is an integer number of samples. The signals $x[n]$ and $w[n]$, of course, need to be lined up properly to take this delay into account before $e[n]$ is actually computed. If the filter order is L then the delay is $\frac{L}{2}$ samples. So the order L should be even i.e. the number of taps $= L + 1$ should be odd. Another issue to be aware of is the passpand ripple of the filter even if the filter has perfectly linear phase. If the signal tone falls where the ripple is not zero, when the input is subtracted from the filter output to compute the error, a poor SNR may result due to this ripple. (If one computed the error spectrum, one would see a big distortion component at the tone frequency due to the ripple). Technically this is a distortion but reflects not the fundamental properties of the A/D converter but the filter passband ripple which is present in most digital filters. Consequently, either the ripple must be small enough so that any distortion produced as a result is far smaller than

the quantization noise being measured or the sinusoid frequency needs to be chosen at a location where the filter has zero ripple. Of course, if the system is not oversampled, no filtering of the output is required. Finally, note that there is no need to perform windowing because time domain measurement implies no implicit periodic extension of signal as is the case when the DFT is computed and so there is no discontinuity to alleviate. Here, we are merely comparing two signals over a given time range, sample by sample.

Until now, we have assumed that the input to the A/D converter system has been a sinusoidal input. One advantage of the time domain technique is that it allows the use of random white noise as input to the A/D converter. Frequency domain techniques will clearly not work for a white input since it will be impossible to separate the noise from the signal in the spectrum. In the time domain, we just need to subtract the input from the output. If the system is oversampled, the A/D converter output will need to be filtered with a linear phase filter and the group delay of this filter as well as any delays due to the A/D converter system should be taken into account in this subtraction. Also, the pass-band ripple should be small enough such that any distortion measured due to it does not eclipse the noise being measured. If the system is not oversampled, no filtering needs to be done at the output of the A/D converter and only delays within the A/D system itself need to be taken into account before subtracting the input from the output.

Appendix C

Integrating P Band MB$\Delta\Sigma$ With M Band QMF Bank

We now show that if a P channel multi-band $\Delta\Sigma$ system is followed by M channel sub-band coder or quadrature mirror filter bank, the computational complexity of the overall structure can be reduced by integrating the digital filters of the two systems. This is demonstrated for the case where the overall signal of the multi-band system is a low-pass signal.

C.0.1 The M Band QMF Bank

First, let us discuss the M channel QMF bank by itself. Fig C.1 shows a block diagram of the QMF bank. The analysis filter in the mth channel is $U_m(z)$ and the synthesis filter in the mth channel is $V_m(z)$. After analysis filtering and downsampling with $W = \exp(-j\frac{2\pi}{M})$, the signal in the mth channel is,

$$\frac{1}{M}\sum_{l=0}^{M-1} X(z^{\frac{1}{M}}W^l)U_m(z^{\frac{1}{M}}W^l) \tag{C.1}$$

After upsampling and synthesis filtering, the signal in the mth channel is,

$$\frac{1}{M}\sum_{l=0}^{M-1} X(zW^l)U_m(zW^l)V_m(z) \tag{C.2}$$

After addition of the channel outputs, the output of the QMF bank is,

$$Y(z) = \sum_{l=0}^{M-1} X(zW^l)\frac{1}{M}\sum_{m=0}^{M-1} U_m(zW^l)V_m(z) \tag{C.3}$$

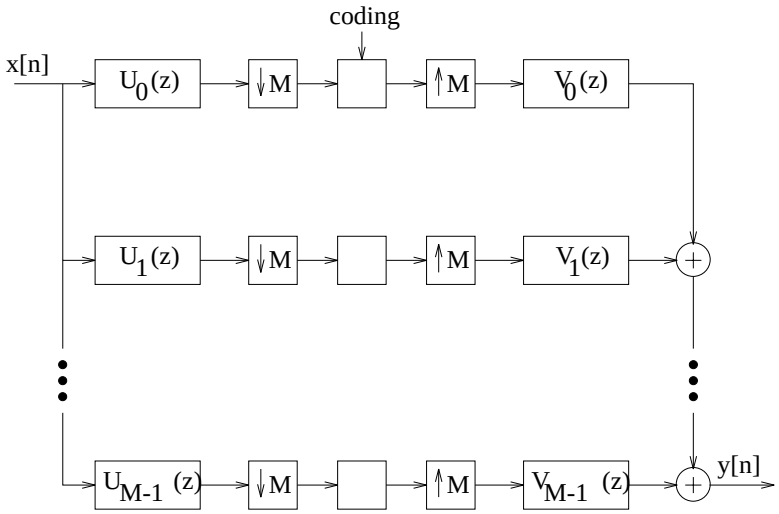


Figure C.1: M channel QMF bank

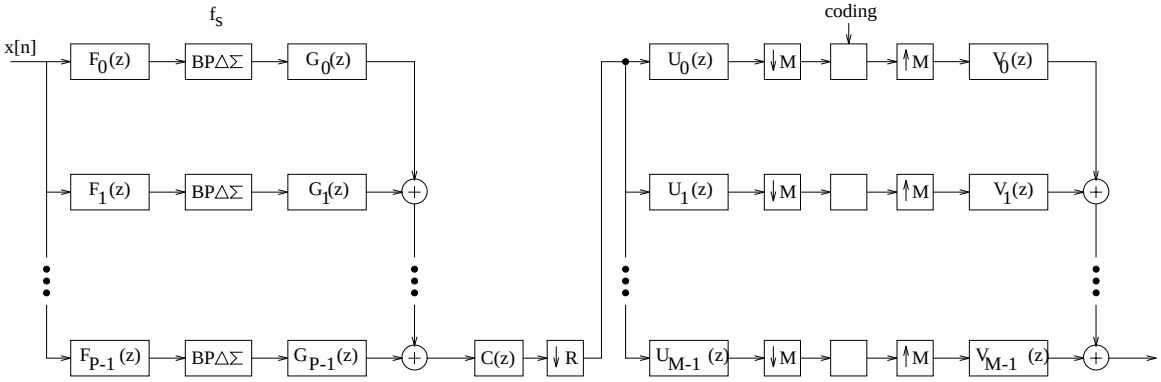


Figure C.2: P channel MBΔΣ followed by M channel QMF bank

Clearly, if the filter bank has the perfect reconstruction property so that $Y(z) = X(z)$, it must be true that, $\frac{1}{M}\sum_{m=0}^{M-1} U_m(zW^l)V_m(z) = \delta[l]$.

C.0.2 P Channel MBΔΣ Followed by M Channel QMF Bank

First, let us append a P channel QMF filter bank after a M channel multi-band ΔΣ converter as shown in Fig C.2. The output of the multi-band ΔΣ after decimation filtering and downsampling by the oversampling ratio R is,

$$\frac{1}{R}\sum_{r=0}^{R-1} X(z^{\frac{1}{R}}W^r)C(z^{\frac{1}{R}}W^r) \tag{C.4}$$

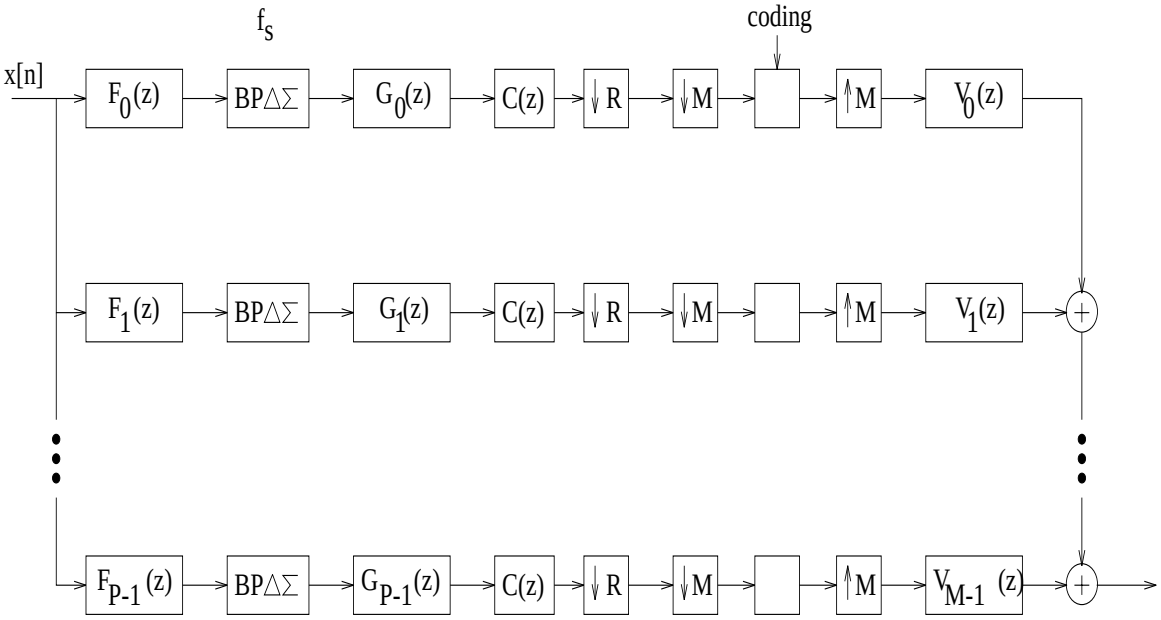


Figure C.3: Integrating P channel MB$\Delta\Sigma$ with M channel QMF bank, $P = M$

where $W = \exp(-j\frac{2\pi}{R})$. If the QMF bank is appended after the downsampler then, if the QMF bank has perfect reconstruction, its output should be the same as in (C.4).

C.0.3 Combining the QMF Bank and MB$\Delta\Sigma$

$P = M$ Case

Let us now consider integrating the QMF bank and the multi-band converter as in Fig C.3 for the case $P = M$. Let the analysis filter of the overall filter bank in the mth channel be $G_m(z)C(z)$. After filtering with $G_m(z)$ and $C(z)$ and then downsampling by R, the signal in the mth channel is

$$\frac{1}{R}\sum_{r=0}^{R-1} X(z^{\frac{1}{R}}W^r)G_m(z^{\frac{1}{R}}W^r)C(z^{\frac{1}{R}}W^r) \tag{C.5}$$

After further downsampling by a factor of M, the signal in the mth channel is,

$$\frac{1}{M}\sum_{l=0}^{M-1}\left[\frac{1}{R}\sum_{r=0}^{R-1} X(z^{\frac{1}{RM}}W^rW'^{\frac{l}{R}})G_m(z^{\frac{1}{RM}}W^rW'^{\frac{l}{R}})C(z^{\frac{1}{RM}}W^rW'^{\frac{l}{R}})\right] \tag{C.6}$$

where $W' = \exp(-j\frac{2\pi}{M})$. After upsampling by a factor of M and synthesis filtering with $V_m(z)$ the signal in the mth channels is,

$$\frac{1}{M}\sum_{l=0}^{M-1}\frac{1}{R}\sum_{r=0}^{R-1}X(z^{\frac{1}{R}}W^rW'^{\frac{l}{R}})C(z^{\frac{1}{R}}W^rW'^{\frac{l}{R}})G_m(z^{\frac{1}{R}}W^rW'^{\frac{l}{R}})V_m(z) \tag{C.7}$$

Finally, after the addition of all the channel responses, the output of the system of Fig C.2 is,

$$Y(z) = \sum_{l=0}^{M-1}\frac{1}{R}\sum_{r=0}^{R-1}X(z^{\frac{1}{R}}W^rW'^{\frac{l}{R}})C(z^{\frac{1}{R}}W^rW'^{\frac{l}{R}})\frac{1}{M}\sum_{m=0}^{M-1}G_m(z^{\frac{1}{R}}W^rW'^{\frac{l}{R}})V_m(z) \tag{C.8}$$

Let us choose,

$$G_m(z) = U_m(z^R) \tag{C.9}$$

Then the term $G_m(z^{\frac{1}{R}}W^rW'^{\frac{l}{R}})$ becomes $U_m(zW^{rR}W'^l)$ which is $U_m(zW'^l)$. Thus we have,

$$Y(z) = \sum_{l=0}^{M-1}\frac{1}{R}\sum_{r=0}^{R-1}X(z^{\frac{1}{R}}W^rW'^{\frac{l}{R}})C(z^{\frac{1}{R}}W^rW'^{\frac{l}{R}})\frac{1}{M}\sum_{m=0}^{M-1}U_m(zW'^l)V_m(z) \tag{C.10}$$

However, if $U_m(z)$ and $V_m(z)$ are chosen as the solutions to a perfect reconstruction QMF bank, then by definition, $\frac{1}{M}\sum_{m=0}^{M-1}U_m(zW'^l)V_m(z) = \delta[l]$ and we have,

$$Y(z) = \frac{1}{R}\sum_{r=0}^{R-1}X(z^{\frac{1}{R}}W^r)C(z^{\frac{1}{R}}W^r) \tag{C.11}$$

This is the same expression as obtained in the case of a QMF bank appended after the multi-band structure followed by decimation filtering and downsampling by R and so the choice of $G_m(z)$ in (C.9) does produce the desired result. Note that the downsampling factor R could have been chosen to be less than R if downsampling to exactly this Nyquist rate was not desired. Of course, the sub-band coder would then be designed to operate at this rate. Also, the design of $G_m(z)$ in (C.9) is similar to that of reduced complexity frequency scaling design based on (9.8).

Computational Complexity

Let us now make a comparison of the computational resources required for the case where the multi-band $\Delta\Sigma$ system is followed by the QMF bank and the case where the two systems

are integrated together as described earlier. The formula (3.64) from [52] is used to estimate the orders of the digital filters required. The order of the filter is given by,

$$L_1 = \frac{sba - 13}{14.6\Delta f} + 1 \simeq \frac{sba - 13}{14.6\Delta f} \tag{C.12}$$

where L_1 is the filter order, sba is the stop-band attenuation, and Δf is the width of the transition band width normalized to a sampling frequency of 1. If sba is roughly the same for all the filters being considered, Δf is the more significant parameter which determines the order. Lumping everything but Δf into a constant K', the order is L_1

$$L_1 = \frac{K'}{\Delta f} \tag{C.13}$$

If Δf is in radians, then

$$L_1 = \frac{2\pi K'}{\Delta f} \tag{C.14}$$

The resources needed for the filter can be broken down into the number of multipliers, adders, and registers or delay elements needed. The number of registers is the same as the filter order while the number of multipliers and adders is equal to the number of non-zero taps in the filter. For $G_p(z)$ filters of Fig C.2 or Fig C.3, since the output of the $\Delta\Sigma$ modulators are assumed to be single bit words, no multiplications are needed and only adders are required. Let us now tabulate the hardware requirements for the multi-band $\Delta\Sigma$ / QMF Bank configurations of Fig C.2 and Fig C.3.

MB$\Delta\Sigma$ followed by QMF bank (Fig C.2)

Let $K = 2tK'$. Then we have the following requirements:

Multipliers and Adders

- For $G_m(z)$, $\Delta f = \left(\frac{2\pi}{P}\right) 2t$ for the calculation of the filter order since $G_m(z)$ are designed from $T_m(z)$ for which $\Delta f \simeq \left(\frac{2\pi}{P}\right) 2t$ and so KP adders are needed. For the P channels, KP^2 adders are required. The adders can operate at a rate of f_s.

- For $C(z)$, $\Delta f = \left(\frac{2\pi}{RP}\right) 2t$, and so $L_1 = KRP$. Therefore, KRP multipliers and adders are required. However, $C(z)$ is followed by downsampling by a factor of R and so can benefit from a polyphase implementation. This means that the multipliers and adders can operate at rate $\frac{f_s}{R}$ instead of f_s.

- For $U_m(z)$, $\Delta f = \left(\frac{2\pi}{P}\right) 2t$, and $L_1 = KP$. Therefore, for the P channels, KP^2 multipliers and adders are required. Since downsampling by a factor of P follows $U_m(z)$, it can also benefits from a polyphase implementation. The multipliers and adders can operate at a rate of $\frac{f_s}{RP}$ instead of $\frac{f_s}{R}$.

Registers (delay elements)

- For $G_m(z)$, each channel needs KRP registers and so the P channels need KRP^2 registers.

- $C(z)$, requires KRP registers. With the benefit of a polyphase implementation these can operate at a rate of $\frac{f_s}{R}$.

- For $U_m(z)$, each channel needs KP registers. So the P channels will require a total of KP^2 registers. With the benefit of the polyphase implementation, the registers can operate at rate $\frac{f_s}{RP}$.

MB$\Delta\Sigma$ Integrated With QMF bank (Fig C.3)

For this arragnement, we have the following requirements:

Multipliers and Adders

- For $G_m(z)$, $\Delta f = \left(\frac{2\pi}{P}\right) 2t$ for the calculation of the filter order since $G_m(z)$ are designed from $U_m(z)$ for which $\Delta f \simeq \left(\frac{2\pi}{P}\right) 2t$ and so $L_1 = KP$. For the P channels, KP^2 adders are required. The adders can operate at a rate of f_s.

- For $C(z)$, $\Delta f = \left(\frac{2\pi}{RP}\right) 2t$, and so $L_1 = KRP$. For the P channels, KRP^2 multipliers and adders are required. $C(z)$ is followed by downsampling by a factor of RP and so with a polyphase implementation, the multipliers and adders can operate at $\frac{f_s}{RP}$.

Registers (delay elements)

- For $G_m(z)$, each channel needs KRP registers and so the P channels need KRP^2 registers.

- Likewise, each $C(z)$ requires KRP registers. For the P channels, a total of KRP^2 registers will be needed. Since each $C(z)$ is followed by downsampling by a factor of RP, with a polyphase implementation, these can operate at rate $\frac{f_s}{RP}$.

Comparing the number of multipliers, adders, and registers required for the two configurations, we observe that the requirements of the separate and integrated multi-band / QMF systems for the $G_p(z)$ filters are the same. For the $C(z)$ filters, the integrated system requires P times the number of adders, multipliers, and registers than the separate systems but the hardware operates at a speed which is P times slower than the hardware in the separate system. Based on area and speed tradeoffs, if we consider these two scenarios for the requirements on $C(z)$ to be equivalent, then on the net we observe that when the $\Delta\Sigma$ filter bank and the QMF bank are integrated, the hardware requirements for the $U_m(z)$ are no longer present. We now examine situations when the number of channels in the multi-band $\Delta\Sigma$ system and the QMF bank are not the same.

$P > M$ Case

This situation may arise when the system quantization SNR requires a large number of $\Delta\Sigma$ channels. If the analysis filters $U_m(z)$ to the perfect reconstruction QMF bank solution have linear phase, we can easily handle this situation. We just need to decompose $U_m(z)$ into a sum of $\frac{P}{M}$ linear phase band-pass filters where $\frac{P}{M}$ should be an integer. This is always possible when $U_m(z)$ has linear phase using an approach similar to the design of the original multi-band converter filters. Thus we will have $U_m(z) = \sum_{l=0}^{\frac{P}{M}-1} U_{m,l}(z)$. We now let $G_p(z) = U_{m,l}(z^R)$ with $m = p \operatorname{div} \frac{P}{M}$ and $l = p \bmod \frac{P}{M}$. To obtain the mth channel, we add groups of $\frac{P}{M}$ channels. $C(z)$ filtering can follow after this addition followed by downsampling by R and then M.

$P < M$ Case

This situation may arise when the system quantization SNR requirement does not require a large number of $\Delta\Sigma$ channels. In this case, the output of each $\Delta\Sigma$ channel need to be expanded into $\frac{M}{P}$ channels to result in M channels. Each of these channels M channels can now be filtered by $G_m(z) = U_m(z^R)$ then by $C(z)$. Downsampling by factors of R and M can now take place.

Appendix D

Effect of Timing Jitter On $\Delta\Sigma$ Modulators

Timing jitter will cause errors in the sampling process. A correctly sampled value of the signal $y(t)$ will become some incorrect value $y_i(t)$ in the presence of jitter. Thus, timing jitter will cause an error in amplitude. The incorrect value $y_i(t)$ can be represented as

$$y_i(t) = y(t) + q(t) \tag{D.1}$$

where $q(t)$ is the error. This error process $q(t)$ in the amplitude domain is a random process and we would like to examine the properties of this process.

If the signal $y(t)$ has slope $y_s(t) = \frac{dy}{dt}$ at time t, the error at time t can be approximated by multiplying the slope with the timing uncertainty or jitter, $j(t)$ at time t. So we have,

$$q(t) \simeq \frac{dy}{dt} j(t) = y_s(t) j(t) \tag{D.2}$$

Let us assume the jitter process $j(t)$ to be a white W.S.S. random process whose proabability distribution function (p.d.f.) is uniformly distributed between JT_s and JT_s where T_s is the sampling clock period and J is some percentage figure which represents the maximum timing jitter.

Let us try to calculate $R_q = \mathcal{E}\{q(t)q(t+\tau)\}$.

$$R_q = \varepsilon\{y_s(t)y_s(t+\tau)j(t)j(t+\tau)\}$$

If we make the rasonable assumption that $y_s(t)$ and $j(t)$ are independent then,

$$R_q = \varepsilon\{y_s(t)y_s(t+\tau)\}\varepsilon\{j(t)j(t+\tau)\}$$

Since $j(t)$ has been assumed to be W.S.S.

$$\varepsilon\{j(t)j(t+\tau)\} = R_j(\tau)$$

Also y_s is obtained from y by the linear process of differentiation. Consequently if $y(t)$ is W.S.S. so is $y_s(t)$ and we have

$$\varepsilon\{y_s(t)y_s(t+\tau)\} = R_{y_s}(\tau)$$

Therefore,

$$R_q = R_{y_s}(\tau)R_j(\tau)$$

Since R_q depends only on τ, it too is W.S.S. and we can find its power spectral density (p.s.d.) $P_q(f)$ as

$$P_q(f) = P_{y_s}(f) \star P_j(f)$$

where $P_{y_s}(f)$ and $P_j(f)$ are the p.s.d. for $y_s(t)$ and $j(t)$ and $\star$ represents linear convolution. For P_{y_s} we have

$$P_{y_s}(f) = P_y(f)|H(f)|^2 = P_y(f)|(j2\pi f)|^2$$

where $H(f)$ represents the frequency domain transfer function of the differentiator. Therefore, we have

$$P_q(f) = \left[P_y(f)(j2\pi f)^2\right] \star P_j(f)$$

Therefore, if the jitter process $j(t)$ is white, the error process, $q(t)$ will also be a white. In the case of ΔΣ modulators, this error process will add directly to the input signal and so benefits from the oversampling but will not benefit from any noise-shaping. The in-band jitter error power is given by:

$$\sigma_q^2 = 2\int_{f_1}^{f_2} \left[P_y(f)(j2\pi f)^2\right] \star P_j(f)df$$

where f_1 and f_2 are the lower and upper band edge frequencies.

Bibliography

[1] S. Renukunta and D. Wells: "Optical memory and blue lasers," *IEEE Potentials*, pp. 14-18, Oct/Nov 1994.

[2] J. Candy and G. Temes: "Oversampling methods for A/D and D/A conversion" in *Oversampling Delta-Sigma Data Converters*, pp. 1-25 (IEEE Press, 1992).

[3] B. Leung: "Theory of $\Sigma - \Delta$ analog to digital converter," *IEEE International Symposium on Circuits and Systems Tutorials,* pp. 196-223, 1994.

[4] P. Aziz, H. Sorensen, J. Van der Spiegel: "An overview of sigma-delta converters: How a 1 bit ADC achieves more than 16 bit resolution," *IEEE Signal Processing Magazine*, pp. 61-83, January, 1996.

[5] A. Oppennheim and R. Schafer: *Discrete Time Signal Processing* (Prentice-Hall, 1989).

[6] W. Bennett, "Spectra of quantized signals," *Bell System Technical Journal*, pp. 446-472, July 1948.

[7] B. Boser and B. Wooley: "The design of sigma-delta modulation analog-to-digital converters," *IEEE Journal of Solid State Circuits*, pp. 1298-1308, December, 1988.

[8] H. Inose and Y. Yasuda: "A unity bit coding method by negative feedback," *Proceedings of the IEEE*, pp. 1524-1535, November, 1963.

[9] J. Candy: "A use of double integration in sigma delta modulation," *IEEE Transactions on Communications*, pp. 249-258, March, 1985.

[10] V. Friedman: "The structure of limit cycles in sigma delta modulation," *IEEE Transactions on Communications*, pp. 972-979, August, 1988.

[11] J. Candy, O. Benjamin: "The structure of quantization noise from sigma-delta modulation," *IEEE Transactions on Communications*, pp. 1316-1323, September, 1981.

[12] S. Hein and A. Zakhor: "On the stability of sigma delta modulators," *IEEE Transactions on Signal Processing,* pp. 2322-2348, July, 1993.

[13] R. Gray: "Spectral analysis of quantization noise in a single-loop sigma-delta modulator with dc input," *IEEE Transactions on Communications*, pp. 588-599, June, 1989.

[14] N. He, F. Kuhlmann, A. Buzo, "Double-loop sigma-delta modulation with dc input," *IEEE Transactions on Communications,* pp. 487-495, April, 1990.

[15] R. Gray: "Quantization noise spectra," *IEEE Transactions on Information Theory*, pp. 1220-1244, November, 1990.

[16] S. Rangan and B. Leung: "Quantization noise spectrum of double-loop sigma-delta converter with sinusoidal input," *IEEE Transactions on Circuits and Systems II,* pp. 168-173, February, 1994.

[17] S. Norsworthy, I. Post, H. Fetterman: "A 14-bit 80kHz sigma-delta A/D converter: modeling, design, and performance evaluation," *IEEE Journal of Solid State Circuits*, pp. 256-266, April, 1989.

[18] S. Norsworthy and D. Rich: "Idle channel tones and dithering in delta sigma modulators," *95th Convention of the Audioengineering Society,*, preprint 3711, October, 1993.

[19] K. Chao, S. Nadeem, W. Lee, and C. Sodini: "A higher order topology for interpolative modulators for oversampling A/D converters," *IEEE Transactions on Circuits and Systems*, pp. 309-318, March, 1990.

[20] P. Ferguson, A. Ganesan, R. Adams: "One bit higher order sigma-delta A/D converters," *Proceedings, IEEE International Symposium on Circuits and Systems*, pp. 890-893, 1990.

[21] F. Op't Eynde, G. Yin, W. Sansen: "A CMOS fourth-order 14b 500k-sample/s sigma-delta ADC converter," *Digest of Technical Papers, International Solid State Circuits Conference*, pp. 62-63, 1991.

[22] R. Adams: "Design aspects of high-order delta-sigma A/D converters," *IEEE International Symposium on Circuits and Systems Tutorials,* pp. 235-259, 1994.

[23] S. Ardalan and J. Paulos: "An analysis of nonlinear behavior in delta-sigma modulators," *IEEE Transactions on Circuits and Systems,* pp. 593-603, June, 1987.

[24] R. Baird and T. Fiez: "Stability analysis of high-order delta-sigma modulation for ADCs," *IEEE Transactions on Circuits and Systems II,* pp. 59-62, January, 1994.

[25] S. Moussavi and B. Leung: "High-order single-stage single-bit oversampling A/D converter stabilized with local feedback loops," *IEEE Transactions on Circuits and Systems II,* pp. 19-25, January, 1994.

[26] D. Anastassiou: "Error diffusion coding for A/D conversion," *IEEE Transactions on Circuits and Systems,* pp. 1175-1186, September, 1989.

[27] R. Schreier and Y. Yang: "Stability tests for single-bit sigma-delta modulators with second-order FIR noise transfer functions," *Proceedings, IEEE International Symposium on Circuits and Systems*, pp. 1316-1319, 1992.

[28] L. Larson, T. Cataltepe, and G. Temes: "Multibit oversampled $\Sigma - \Delta$ A/D convertor with digital error correction," *Electronics Letters*, pp. 1051-1052, August 4, 1988.

[29] M. Sarhang-Nejad and G. Temes: "A high-resolution multibit $\Sigma\Delta$ ADC with digital correction and relaxed amplifier requirements," *IEEE Journal of Solid State Circuits*, pp. 648-660, June, 1993.

[30] T. Leslie and B. Singh: "An improved sigma-delta modulator architecture," *Proceedings, IEEE International Symposium on Circuits and Systems*, pp. 372-375, 1990.

[31] A. Hairapetian, G. Temes, and Z. Zhang: "Multibit sigma-delta modulator with reduced sensitivity to DAC nonlinearity," *Electronics Letters*, pp. 990-991, May 23, 1991.

[32] L. Carley: "A noise-shaping coder topology for 15+ bit converters," *IEEE Journal of Solid State Circuits*, pp. 267-273, April, 1989.

[33] B. Leung and S. Sutarja: "Multibit sigma-delta A/D converter incorporating a novel class of dynamic element matching technique," *IEEE Transactions on Circuits and Systems II,* pp. 35-51, January, 1992.

[34] F. Chen and B. Leung: "A high resolution multi-bit sigma-delta modulator with individual level averaging," *Digest of Technical Papers, IEEE Symposium on VLSI Circuits*, pp. 101-102, June, 1994.

[35] Y. Matsuya, K. Uchimura et al: "A 16-bit oversampling A-to-D conversion technology using triple-integration noise shaping," *IEEE Journal of Solid State Circuits*, pp. 921-929, December, 1987.

[36] M. Rebeschini, N. van Bavel, P. Rakers, et al: "A 16-b 160-kHz CMOS A/D converter using sigma-delta modulation," *IEEE Journal of Solid State Circuits*, pp. 431-440, April, 1990.

[37] D. Ribner: "A comparison of modulator networks for high-order oversampled $\Sigma\Delta$ analog-to-digital converters," *IEEE Transactions on Circuits and Systems*, pp. 145-159, February, 1991.

[38] K. Uchimura, T. Hayashi, T. Kimura and A. Iwata: "Oversampling A-to-D and D-to-A converters with multistage noise shaping modulators," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, pp. 1899-1905, December, 1988.

[39] L. Longo and M. Copeland: "A 13 bit ISDN-band oversampled ADC using two-stage third order noise shaping," *Proceedings, IEEE Custom Integrated Circuits Conference*, pp. 21.2.1-21.2.4, 1988.

[40] G. Yin, F. Stubbe, W. Sansen: "A 16-b 320-kHz CMOS A/D converter using two-stage third-order $\Sigma\Delta$ noise shaping," *IEEE Journal of Solid State Circuits*, pp. 640-647, June, 1993.

[41] L. Williams and B. Wooley: "Third-order sigma-delta modulator with extended dynamic range," *IEEE Journal of Solid State Circuits*, pp. 193-202, March, 1994.

[42] P. Benabes, A. Gauthier, D. Billet: "A new wideband sigma-delta convertor," *Electronics Letters*, pp. 1575-1576, August 19, 1993.

[43] R. Schreier and M. Snelgrove: "Bandpass sigma-delta modulation," *Electronics Letters*, pp. 1560-1561, November 9, 1989.

[44] S. Jantzi, R. Schreier and M. Snelgrove: "Bandpass sigma-delta analog-to-digital conversion," *IEEE Transactions On Circuits and Systems,* pp. 1406-1409, November, 1991.

[45] L. Longo and B. Horng: "A 15b 30kHz bandpass sigma delta modulator," *Digest of Technical Papers, International Solid State Circuits Conference*, pp. 226-227, 1993.

[46] S. Jantzi, M. Snelgrove, P. Ferguson: "A 4th-order bandpass sigma-delta modulator," *IEEE Journal of Solid State Circuits*, pp. 282-291, March, 1993.

[47] G. Tröster, H. Dreβler, et al: "An interpolative bandpass converter on a 1.2μm BiCMOS analog/digital array," *IEEE Journal of Solid State Circuits*, pp. 471-477, April, 1993,

[48] D. Ribner: "Multistage bandpass delta sigma modulators," *IEEE Transactions on Circuits and Systems II*, pp. 402-405, June, 1994.

[49] R. Crochiere and L. Rabiner: "Interpolation and decimation - a tutorial review," *Proceedings of the IEEE*, pp. 300-331, March, 1981.

[50] J. Candy: "Decimation for sigma delta modulation," *IEEE Transactions on Communications*, pp. 72-76, January, 1986.

[51] S. Chu and C. Burrus: "Multirate filter design using comb filters," *IEEE Transactions on Circuits and Systems*, pp. 913-924, November, 1984.

[52] T. Parks and C. Burrus: *Digital Filter Design,* Wiley, 1987.

[53] E. Dijkstra, M. Degrauwe, J. Rijmenants, O. Nys: "A design methodology for decimation filters in sigma delta A/D converters," *Proceedings, IEEE Symposium on Circuits and Systems,* pp. 479-482, 1987.

[54] B. Leung: "Design methodology of decimation filters for oversampled ADCs based on quadratic programming," *IEEE Transactions on Circuits and Systems,* pp. 1121-1132, October, 1991.

[55] Z. Ma and B. Leung: "Polyphase IIR decimation filter design for oversampled A/D converters with approximately linear phase," *IEEE Transactions on Circuits and Systems II,* pp. 497-505, August, 1992.

[56] S. Lewis: "Video rate analog to digital conversion using pipelined architectures,", Ph.D Thesis, University of California, Berkeley, 1987.

[57] A. Petraglia and S. Mitra: "High speed A/D conversion using QMF banks," *Proceedings, IEEE International Symposium on Circuits and Systems*, pp. 2797-2800, 1990.

[58] W. Black and D. Hodges: "Time interleaved converter arrays," *IEEE Journal of Solid State Circuits*, pp. 1022-1029, December, 1980.

[59] P. Vaidyanathan: "Multirate systems and filter banks," Prentice-Hall, p. 406, 1993.

[60] V. Sathe and P. Vaidyanathan: "Effects of multirate systems on the statistical properties of random signals," *IEEE Transactions on Signal Processing*, pp. 131-146, January, 1993.

[61] P. Vaidyanathan and P. Hoang: "Lattice structures for optimal design and robust implementation of two-channel perfect reconstruction QMF banks," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* pp. 81-94, January, 1988.

[62] P. Aziz in Electrical Engineering research meeting with K. Laker, H. Sorensen, J. Van der Spiegel, University of Pennsylvania, February, 1994.

[63] S. Jantzi, K. Martin, M. Snelgrove, A. Sedra: "Complex bandpass $\Sigma\Delta$ converter for digital radio," *Proceedings, IEEE International Symposium on Circuits and Systems*, pp. 453-456, 1994.

[64] P. Aziz, H. Sorensen, J. Van der Spiegel: "Performance of complex noise transfer functions in bandpass and multi band sigma delta systems," *Proceedings, IEEE International Symposium on Circuits and Systems*, pp. 641-644, 1995.

[65] R. Schreier: "An empirical study of high-order single-bit delta-sigma modulators," *IEEE Transactions on Circuits and Systems II*, pp. 461-466, August, 1993.

[66] S. Jantzi, C. Ouslis, A. Sedra: "Transfer function design for $\Delta\Sigma$ converters," *Proceedings, IEEE International Symposium on Circuits and Systems*, pp. 433-436, 1994.

[67] C. Ouslis, A. Sedra: "Designing custom filters: From specification to realization in the unique filtorX design environment," *IEEE Circuits and Devices Magazine*, pp. 29-37, May, 1995.

[68] P. Aziz, H. Sorensen, J. Van der Spiegel: "Theoretical and practical performance of real and complex delta-sigma band-pass noise transfer functions," *in preparation for submission to IEEE Transactions on Circuits and Systems II.*

[69] K. Nagaraj, K. Singhal, T.R. Viswanathan, J. Vlach: "Reduction of finite-gain effect in switched-capacitor filters," *Electronics Letters*, July 18, 1985, pp. 644-645.

[70] K. Nagaraj, T.R. Viswanathan, K. Singhal, J. Vlach: "Switched-capacitor circuits with reduced sensitivity to amplifier gain," *IEEE Transactions on Circuits and Systems*, May, 1987, pp. 571-574.

[71] P.W. Bosshart: "A multiplexed switched capacitor filter bank," *IEEE Journal of Solid State Circuits*, December, 1980, pp. 939-45.

[72] K.R. Laker, P.E. Fleischer, A. Ganesan: "Parasitic insensitive, biphase switched capacitor filters realized with one operational amplifier per pole pair," *Bell System Technical Journal*, May-June, 1982, pp. 685-707.

[73] Q. Liu, M. Snelgrove, A. Sedra: "Switched capacitor implementation of complex filters," *Proceedings, IEEE International Symposium on Circuits and Systems*, 1986, pp. 1121-1124.

[74] P.E. Fleischer, A. Ganesan, K.R. Laker: "Parasitic compensated switched capacitor circuits," *Electronics Letters*, November 26, 1981, pp. 929-931.

[75] R. Khoini-Poorfard and D. Johns: "Time-interleaved oversampling convertors," *Electronics Letters*, pp. 1673-1674, September 16, 1993.

[76] P. Aziz, H. Sorensen, J. Van der Spiegel: "Multiband sigma-delta modulation," *Electronics Letters*, pp. 760-762, April 29, 1993.

[77] P. Aziz, H. Sorensen, J. Van der Spiegel: "Multiband sigma delta analog to digital conversion," *Proceedings, IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 249-252, 1994.

[78] I. Galton and H. Jensen: "Delta-sigma modulator based A/D conversion without oversampling," *IEEE Transactions on Circuits and Systems*, December, 1995.

[79] B. Song: "A 4th order bandpass delta-sigma modulator with reduced number of op amps," *Digest of Technical Papers, International Solid State Circuits Conference*, 1995.

[80] Y.C. Lim: "Frequency-response masking approach for the synthesis of sharp linear phase digital filters," *IEEE Transactions on Circuits & Systems*, pp. 357-364, April 1986.

[81] J.H. McClellan, T.W. Parks, L.R. Rabiner, "FIR linear phase filter design program," in *Programs for Digital Signal Processing*, IEEE Press, 1979.

[82] R.E. Crochiere and L.R. Rabiner: *Multirate digital signal processing*, Prentice-Hall, 1983.

[83] P. Aziz, H. Sorensen, J. Van der Spiegel: "Frequency sliced delta-sigma A/D converters," *in preparation for submission to IEEE Transactions on Circuits and Systems II.*

[84] R. Khoini-Poorfard, L. Lim, D. Johns: "Time-interleaved oversampling A/D converters: theory and practice," *To appear in IEEE Transactions on Circuits and Systems II.*

[85] Dr. Palaksha Setty, Member of Technical Staff, *Lucent Bell Laboratories*, private communication.

[86] M.J. Demler: *High-Speed Analog-to-Digital Conversion*, Academic Press, 1991.

[87] F.J. Harris: "On the use of windows for harmonic analysis with the discrete-fourier transform," *Proceedings of the IEEE*, January, 1978.

[88] L.B. Jackson: *Digital Filters and Signal Processing*, Kluwer Academic Publishers, 1986.

Index